

South Africa gets out of hand

The present sad condition of the Republic of South Africa is unlikely to last for very long. But the timescale of general impatience may be dangerously short.

WHAT can be done and what should be said about the growing crisis in South Africa? *Nature* has no competence in the matter, which lies outside the terms of reference implicitly defined by a readership of professional scientists. Of course, there are some aspects of the condition of South Africa on which science has a bearing. There can be no genetic basis for the inequitable and offensive system of apartheid, which is not to say that apparent differences in the attainment of different races may not be induced by the systematic social deprivation of one (or several) of them. It is also pertinent, as in the argument begun just under a year ago by the thoughtless Southampton organizers of September's World Archaeological Congress, that academics outside South Africa cannot ostracize fellow-intellectuals working at South African institutions without irreparably and permanently damaging the cause of international scholarship. But there are also some commonsense considerations that deserve more attention than they are being given.

Those outside South Africa wishing to encourage change should prudently begin with an assessment of the time over which their hopes are likely to be realized. The temptation is to miscalculate that the system of apartheid is so near collapse that only modest pressure from outside will send it tumbling. This is why the cry for general economic sanctions is now widespread, not merely from outside South Africa, but by many of the black leaders within the republic. But that underestimates the capacity for survival of the Nationalist South African government. One of the most distasteful features of the apartheid system is that many of the black workers on whom the economy depends belong titularly not where they work but to artificial living places many miles away, to which they are supposed to return when they are out of work. If sanctions were effective enough to hurt the economy of South Africa, the republic government would be acting within its own laws if it chose to send the unemployed off to their "homelands" to starve in obscurity. The process would be difficult but not impossible, and would buy more time. Is it possible that well-wishers outside South Africa have allowed their distaste of the present system to lead them to too sanguine a view of the weakness of a government whose determination is, in other connections, its hallmark?

But if not sanctions, what? That is the general dilemma. The simplest answer is "selective sanctions", by which is meant a variety of schemes ranging from the cancellation of permission for South African aircraft to land elsewhere to a collective decision not to buy South African exports of certain kinds, fruit for example. But if full sanctions would not produce the intended effect, can half-sanctions succeed? (A few symbolic actions might nevertheless be worthwhile.) The difficulty running through these tortuous considerations is the frustration of governments elsewhere that they have no obvious means except the passive device of sanctions legally to influence internal affairs in South Africa.

Luckily, that does not imply that nothing can be done. Although the international banking system has been quietly helping to bring about the "disinvestment" in South Africa for which well-wishers have been asking, would it not now make sense, as

a condition for the continued operation of South African companies as subsidiaries of companies registered elsewhere, that they should promptly follow employment practices of the kind that would be expected of them outside South Africa, treating people of different races on an equal footing and providing for the disadvantaged the kind of technical training that would allow them to function properly in their jobs? The cost could be high, but not necessarily greater than that of sanctions and certainly less than the cost of the international chaos that will supervene if South Africa falls apart. Indeed, it might even be worthwhile to encourage the South African subsidiaries of overseas corporations to break the apartheid laws, as Barclays Bank has just done by lending money to a black employee to buy a house in a district reserved for whites. When industrialized governments do not shrink from telling companies to instruct their overseas subsidiaries to which customers they may and may not sell high technology, why should they shrink from using this power constructively within South Africa? □

Trouble on the farm

The problems now facing farmers will not go away overnight.

MOST developed countries have long since been forced by economic reality to reduce the scale on which they maintain once-traditional industries such as shipbuilding and steel manufacture. In many places, farming will be the next in line. So much has been clear for several years, some would say decades. Much of the food that farmers in developed countries now grow with the crucial help of government subsidies could already be grown more cheaply elsewhere, in Australia, New Zealand, Argentina and Zimbabwe, for example. As time passes, the list of potential suppliers of temperate foodstuffs will be or could be increased by the addition of many now-developing countries, which would conform well with Adam Smith's principle that the economic use of resources requires a division of labour: in the modern world, some must grow food and others make computers.

That is what the textbooks say. The reality is very different. The subsidies industrialized governments pay their farmers, which have ensured over the years that the industrialized world is awash with food that impoverished nations cannot afford to buy, serve a multitude of functions, not all of them respectable. Occasionally, the old strategic argument still surfaces that a country less than self-sufficient in food production is vulnerable to starvation in time of war, but this can hardly carry weight when most governments fear that future wars will be over before most people know they have begun, and when government warehouses are in any case filled to the brim with food that nobody wants. The social argument, that the farming community is the repository of an invaluable tradition whose erosion would spell the loss of crucial elements in the constitution of advanced societies, contrasts oddly with the continual war between farmers and many of the societies in which they are embedded over environmental issues, for example. (Too often,



farmers cut the figure of state-licensed vandals.) That farming is a social problem is, on the other hand, undeniable: the plight of farmers (and their banks) in the Middle West of the United States during the past two years is one vivid proof. The nub of the difficulty is that most industrialized governments, after decades of encouragement of increased farm production, now find themselves morally committed to a large group of people whose lives have been spent in farming, who have no wish to do anything else and who are not outstandingly fitted to do other things. Governments also recognize, of course, that farmers are also voters, often noisy.

This dilemma is nicely illustrated by the state of British farming. Last year was in many ways disastrous. The government's *Annual Review of Agriculture 1986* (HMSO, Cmnd 9708), published at the beginning of the year, showed that as a consequence of the appalling weather, farm income fell in 1985 by 43 per cent compared with the previous year, which had been a good year. In a year when there were 240,000 farms in Britain (a steadily declining number), farmers and their spouses were left with a cash income of merely £1.154 million, just about half the total scale of subsidy for British farmers last year (£2,205 million) and even less than the proportion of the farming subsidy (£1,308 million) provided by the Common Agricultural Policy by means of which European farming is sustained at an uneconomically high level.

In 1985, in other words, it would have been cheaper for the British taxpayer merely to have provided farmers with the cash they are supposed to have earned from their farming operations and to have bought in their production from the huge stocks of food the Community has built up over the years. Yet farmers are plainly undismayed by this alarming calculation. Farmers spent more on new machinery and building than the cash that stayed in their pockets, while the price of farm land remained buoyant, as the real estate dealers say.

Can this state of affairs be allowed to continue? Common sense says no. Now even the British government seems to be having second thoughts. In the past few weeks, there have been repeated hints that the scale of traditional farming must decline. One common suggestion is that farmers must take to growing other things than the traditional crops, and that British farmers should in particular take to growing trees. The prospect is not cheerful. So long as the incentive is financial, the trees that farmers grow will be cheap and nasty (see *Nature* 322, 101; 1986) like those favoured by the Forestry Commission. Environmental conflicts are more likely to be sharpened than diminished in the process. But trees grow slowly, and will not provide the continuing employment that farmers have come to regard as their birthright, even though the taxation system could no doubt be modified in such a way as to put £1,000 million plus in their pockets every year. Why not instead bite the unpalatable bullet, and draw up a long-term strategy for the industry that will properly quantify the supposed non-economic benefits, set these off against the equally intangible but important disbenefits of enforced rural life, and declare that the time will have to come, some years from now, when farming will be run on economic lines? □

Shutting barn doors

The consequences of the Chernobyl accident are about to become diplomatic, thanks to IAEA.

WHEN the Soviet reactor at Chernobyl, north of Kiev, went out of control on 26 April, it quickly became apparent, in Sweden where the fallout was first detected and, then, elsewhere in Western Europe, that the formal arrangements for dealing with trans-frontier radioactive pollution are inadequate. At the outset, there was no obvious means by which the Soviet Union could be required to provide fuller information about the character and the scale of the release of radioactive material

from the Chernobyl reactor. (It is now clear that the Soviet officials in charge of the reactor were not themselves aware of what had gone wrong, and that they may even have concealed what they knew about the scale of the accident from Moscow.) Inevitably, the Soviet Union was subjected to a great deal of verbal drubbing from Western governments, many of them eager to complain that Soviet secretiveness, a long-standing nuisance, had become a danger to other people's lives. Yet, curiously, it has been overlooked that, for the past three years, the International Atomic Energy Agency (IAEA) at Vienna has been nursing a kind of international agreement on trans-frontier radioactive pollution. At a conference opening at Vienna later this month, the agency is hoping to turn its draft agreement into an agreement that will stick.

Before this laudable exercise is written off as an attempt to shut the stable (or barn) door after the horse has bolted, some consideration should be given to the complexity of the issues that will arise during the coming weeks. The starting-point for IAEA's conference is the set of recommendations produced in April 1984 on the management of nuclear accidents that might cause fallout beyond the boundaries of the states housing faulty reactors and other nuclear facilities. Sensibly, indeed inevitably, the experts concluded that there should be an element of planning in advance.

Each reactor operator should have a well-defined set of criteria for telling when an accident must be taken seriously, must tell the authorities in neighbouring states about these criteria and must then identify, also in advance, the nuclear installations that are potentially a source of international trouble. All that makes sense, as does the requirement that there should be identified communications links by which news of accidents may be signalled if and when they happen. What IAEA now seeks, in the chastened atmosphere after Chernobyl, is an agreement between its member states that these modest guidelines will be adopted. Whether it will succeed is another matter.

Here are some of the impediments to agreement at Vienna. In many places, the exact design of nuclear installations and even their intended function is still a secret, for either military or quasi-military reasons. The Soviet Union, which declares the nature of its operating reactors to IAEA as a matter of course, says little about its other facilities such as separation plants. France is similarly coy about the division between the civil and military aspects of its nuclear plants.

Other states such as India and Israel (both members of IAEA but non-signatories of the Non-Proliferation Treaty) think they have an interest in keeping their neighbours guessing about their nuclear programmes, believing that their own security will be enhanced if they can give the impression of being capable of making nuclear weapons. Such states will not readily fall in with the suggestion that they should disclose the nature of their nuclear installations.

But even elsewhere, where a state has nothing (or nothing much) to hide, there will be a profound reluctance to say in advance that this or that nuclear plant is in principle capable of going wrong, and of causing trouble to others. Indeed, it is easy to imagine how information of that kind could be used malevolently by those who believe that Chernobyl gives them a licence to halt the nuclear industry in its tracks.

So this month's meeting at Vienna is bound to be a tense affair. Even enlightened governments will be uneasy about the disclosures they will be asked to make. Yet their long-term interest is precisely the kind of understanding on the management of major nuclear accidents that IAEA's guidelines would set in place. If one consequence will be the abatement of senseless secrecy in the management of nuclear installations, that will be an uncovenanted benefit of what is, in any case, an essential measure. And the fear that too much openness will put too much ammunition in the hands of the anti-nuclear people is unreal; part of the trouble now is that too many authorities have been too secretive for too long. □

European research

Framework programme poses political questions

It began to emerge clearly last week that the scale of the European Commission's new "framework programme" of cooperative research and development in Europe is so large that a phase of real interaction between international and national research policies in Europe is now beginning, with national programmes the likely "losers" as research becomes more international.

This internationalization of science seems to be becoming a particular possibility in Britain, where technology minister, Mr Geoffrey Pattie, said last week that he had "no trouble" with the content of the European Commission's 1987-91 "framework programme", the definitive version of which the Commission should place before member states in the next few weeks.

Pattie's support, though qualified when it comes to matters of scale, is significant because, in line with Britain's current presidency of the European council of ministers, Pattie will preside for the next six months over the research council. And Britain attaches "considerable importance" to concluding political negotiations over the content and financing of the programme by the end of this year, Pattie says.

Pattie is concerned, however, at the possible cost of the programme (earlier this year, the Commission was seeking a near tripling of present spending from the present 4,300 million European Currency Units); he demands more critical pre-assessment and analysis of Commission projects and says a final decision on funding may go to cabinet level.

In fact, Pattie's concern reflects the wider issue: that the levels of research spending recommended by the Commission are at last reaching a politically noticeable level. From around 2 per cent of member states' net research and development budgets, the Commission has moved in its framework programme to nearer 4-6 per cent. In individual projects, the levels are even higher: in materials science, for example, where European national spending lags far behind that of the United States and Japan, initial Commission spending proposals in the framework programme reached a fifth of all European research. This seemed to be going well beyond the mere "catalysis" of European collaboration for which the European Commission has previously argued; and at these levels it has begun to be difficult even for enthusiasts of European collaboration — like Mr Pattie — to argue that European research programmes should

not substitute for national research, but add to it. Increases in European programmes are in danger, in Britain at least, of causing reductions in corresponding national ones.

Thus, in money-conscious Britain, where many of the financial issues are clearest, Treasury sources stress that ministries benefiting from any substantial increase in Commission research funding — by winning grants — will have to account for these receipts, and may have their national support commensurately reduced. This near-automatic system of "attribution" of Commission receipts could affect a number of national programmes, not least because British researchers appear to be particularly effective at winning Commission support: last year, Britain won 27 per cent of available Commission research funding, although the country pays only 20 per cent of the Commission's total budget. In research, Britain's receipts exceed "*juste retour*".

Paradoxically, however, this success of British scientists in winning funds from Brussels could in future be penalized by Treasury insistence on the attribution of receipts. Increases under the framework programme could thus conflict with the research of the Department of Energy, and particularly of the Department of Trade and Industry. The Commission's plan to boost its pilot-scale "stimulation programme" — which has successfully combined basic research groups in joint research throughout Europe — could result, under the Treasury attribution rules, in a commensurate decrease in the funds of the research councils, or the universities directly. Sir David Phillips, who is both chairman of the Advisory Board for the Research Councils (which distributes basic research funding in Britain) and a member of the CODEST scientific committee which advises the Commission on science policy, admits that "a key question" now is how far, and whether, national budgets should be cut to support European cooperation. This is why Pattie considers that finding the money for the framework programme may become an issue for the whole government.

The Treasury itself remains unconvinced by the Commission's arguments for an increase in research, and opposes Pattie's view that the framework should not substitute for national programmes. Pattie himself feels he unfortunately gets little help from the Commission — which however important the programmes proposed tends to argue by exhortation.

More detailed and precise arguments are necessary, Pattie says. The framework programme itself should therefore include an important element of assessment to ensure and detail the "effectiveness" and "relevance" of projects and to sharpen arguments for ministers.

Despite the drive for "effectiveness" of the framework programme, however, "flat out market considerations cannot be the order of the day", Pattie insists, because the research-oriented framework programme must take into account the needs of the smaller countries in Europe. These have felt neglected by the quite separate (non-Commission) Eureka technology programme, in which market forces are to the fore. But framework research should have a long-term relevance to European technological competitiveness. An increase in industrial participation, on a sliding scale, as Commission research projects developed, would sharpen Commission programmes Pattie suggested. It would help "to keep testing out the project on industry" and to "make people shape up".

Robert Walgate

German universities

Hamburg

In the past fifteen years, federal and *Länder* (regional) governments have spent DM38,000 million to create an extra 300,000 places at West German universities. But the era of expansion is now near its end, according to the commission for university building.

In its sixteenth framework plan, published last week, the commission recommends that a further DM9,000 million be spent between now and 1990. High priority is given to the extension of natural science and engineering faculties, among them a new faculty of informatics in Karlsruhe, the next phase of the new Biozentrum in Würzburg and a new centre for interdisciplinary research in Frankfurt. The greater part of the money, however, will go for medical institutions and clinics.

The Federal Minister of Education, Dorothee Wilms, emphasized that the goal of 850,000 university places had not yet been reached, but would be brought close with the creation of another 20,000 places in the near future. Although there are now signs of a downward trend in the number of students entering university, there are still around 1.3 million students, far more than there are places. Students have grown accustomed to overcrowded lecture theatres without room even to sit on the floor and inadequate library and laboratory space. As the deputy commission head, Anke Brunn, pointed out, there is no sign of the feared "education catastrophe", the collapse of education due to excess student numbers.

Jürgen Neffe

Spain

More university reforms needed

Barcelona

THERE are said to be two national spectacles in Spain, bullfights and "oposiciones", the recruiting system for the civil service. It is a matter of discussion which one is the bloodier. Reforms in access to university and CSIC (Spanish Science Research Council) positions have modified the "oposiciones", but recent experience suggests that some of the most controversial elements of the system still exist.

Spanish professors and researchers in public institutions are considered as civil servants and therefore have permanent tenure, which tends to create lethargy in many Spanish institutions. That was certainly the opinion, before coming to power, of many members of the Socialist Party, now starting its second four-year term in Spanish central government. Recent regulations have led to invited professors and research staff recruited on short-term contracts. The reforms introduced in universities and in CSIC in the past four years have, however, maintained the traditional civil service system and the selection by "oposiciones".

The spectacle itself is much less colourful than it used to be. The "oposiciones", in the nineteenth-century tradition but modified in the 1930s during the Second Spanish Republic, used to be a series of public examinations, theoretical or practical, designed to choose the ablest candidate for the job. Now, a candidate has to pass only two examinations, one on his personal curriculum and the second on a subject related to his field of research. A committee formed for every position or groups of similar positions makes its decision by voting and without appeal. The committee can put any kind of questions to the candidate. In some cases, the experience may be traumatic for candidates, committees (which may sit for weeks before reaching a decision) and for departments.

The Law for University Reform has introduced two important modifications. One is that universities now have the right to appoint two of the five-member committee that selects the candidates; the other three members are chosen by a computer at random from the list of professors of the same field.

The second change is that university departments may propose a "profile" of the candidate. The experience of the past two years shows that the result has normally been to slant appointments to specific people already working in the same department. The actual result is an increase in in-breeding in university departments and many people are worried about the long-term consequences. Hundreds of new positions are being advertised as a

consequence of the Law of University Reform; for example, the Autonomous University of Barcelona has had nearly 300 new positions in 1986 alone. But the rules may be modified in the near future in the light of experience.

In CSIC, the system shows some important differences. On the one hand, positions are advertised once a year and relate to specific institutes and specific subjects; on the other hand, since last year, all committee members are appointed by the president of CSIC and often include members of university departments as well as CSIC staff. In each of the past two years, about 200 positions have been advertised, after a long time without openings. Last year around 400 candidates applied for the positions. In some fields such as molecular biology, there were four candidates for each position,

which means that in many cases only one candidate applied, or none. This result has been interpreted as showing the need for a long-term policy for recruiting research personnel and for attracting scientists now working abroad. New statutes for CSIC are expected to be issued before the end of this year that will include a different structure for research staff.

The "oposiciones" was a system designed more than a hundred years ago to make Spanish public administration more accessible. It has been widely recognized that a new system would be needed before the end of the twentieth century to select the appropriate people for research or universities. But the rules of Spanish administration are difficult to change, and the recent regulations have tried to adapt an obsolete system to modern Spanish society.

It is generally felt in universities and CSIC that further steps are needed in any case to achieve a flexible and efficient recruiting system for Spanish higher education and research. **Pedro Puigdoménech**

India

Space programme uncertainty

New Delhi

INDIA is still not sure when or how to get its multipurpose Insat 1-C satellite launched into geostationary orbit to provide continuity in its space-based meteorological, television and telecommunication services.

Under a contract with the US National Aeronautics and Space Administration (NASA), the satellite, along with an Indian payload specialist, was to have been carried by the space shuttle this month. The indefinite postponement of shuttle flights after the Challenger explosion, and uncertainty about the availability of the alternative Delta rockets, has upset the schedule of the Indian Department of Space (DOS).

According to DOS secretary Professor U.R. Rao, negotiations are now going on with the European Space Agency for Insat 1-C to be accommodated in one of the Ariane launches before the end of 1987. With Ariane solidly booked for the next two years, and with its own schedule upset by the failure of the Ariane-IV launch in March, it is likely that Insat 1-C may not leave the ground before 1988. India's only functioning satellite, Insat 1-B, has been taking the full work load since August 1983 after its companion, Insat 1-A, was disabled by on-board power failure. The US Ford Aerospace Corporation, which built the three satellites for India, has been asked to deliver a fourth, Insat 1-D, by 1988.

Continuity of the satellite service is important, for India now relies significantly on Insat 1-B for weather forecast, television broadcast and telephone links be-

tween remote places. The satellite provides 4,000 two-way telephone circuits and has helped to network India's 173 television and 91 radio stations, as well as beaming television programmes directly to 2,000 community receivers in villages and relaying educational programmes to 1,953 schools and 62 colleges.

Insat 1-B is the only geostationary satellite over the Indian Ocean providing meteorological images to the world, and duplicate Insat weather data are supplied daily to the United States. Apart from ensuring continuity of services, India is anxious to get Insat 1-C launched as soon as possible to avoid possible political complications arising out of delay in occupying the allotted slot in the geostationary orbit. India will continue to rely on foreign launchers, as it is unlikely to be able itself to put satellites into geosynchronous orbit before the year 2000.

Insat 1-C is not the only project beset by delay. The test flight of the Augmented Satellite Launch Vehicle (ASLV), designed to lift a 150-kg payload, was scheduled for last year but has yet to take place. The launch of an Indian-built remote-sensing satellite from the Soviet Union, scheduled for September this year, has been postponed. India's only operational rocket, SLV-3, has not flown for the past two years.

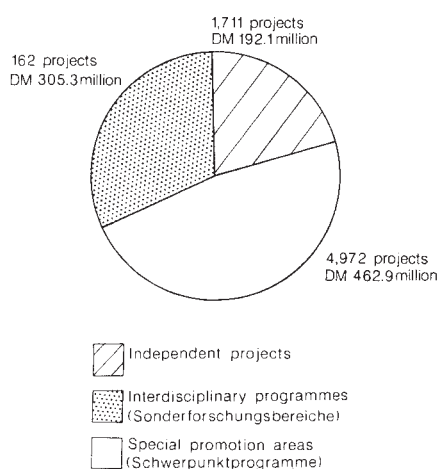
According to DOS, the destruction of major facilities at its east coast Sriharikota launch complex is one reason for the slowdown of space activities. The complex was damaged in a cyclone two years ago and plans are afoot to shift it to a safer location. **K.S. Jayaraman**

The deutschmarks that count

Hamburg

THE annual meeting of the Deutsche Forschungsgemeinschaft (DFG), the West German Science Foundation, was held in Bonn last week. DFG's new president, Hubert Markl, pointed out that all DFG's activities were met with a budget of less than DM 1,000 million, just two per cent of the nation's research and development expenditure and, more significantly, only 60 per cent of the amount spent to subsidize the public theatres.

In 1985, DFG spent DM 982.6 million for the support of research and researchers, with 74 per cent going on personnel costs. Most of the money came from the federal government (DM 578.6 million) and the *Länder* (DM 397.4 million); expenditure was divided as below.



Markl rejected criticisms by the Wissenschaftsrat, the senior science advisory board, which had earlier said that doctoral education in West Germany is both too long and ineffective. Some 4,000 doctoral students are supported by DFG. Doctoral research is carried out in well controlled projects of quality, Markl said. He singled out for praise the contribution made by female scientists.

Chancellor Helmut Kohl made a speech in which he said that "quality and performance have reached a higher level", partly due to a retreat from the politicization of science. Despite increasing debate on the dangers of nuclear power and of science in general following the Chernobyl disaster, Kohl demands that politics should be kept out of science.

Hans-Werner Franke, the president of the *Länder* Ministers of Culture, Education and Church Affairs (Kultusministerkonferenz), demanded that the existing university system should be consolidated even with a decreasing student population. That may create a chance to switch some resources to research and to much-needed reforms.

Jürgen Neffe

US technology

Making rivals work together

Washington

FROM the outset, the US Microelectronics and Computer Technology Corporation (MCC) was risky business. The plan to bring together arch-competitors in the semiconductor and computer industries in a cooperative research institute had no precedent, and even the Justice Department considered challenging the consortium's formation. After operations began in 1983, MCC president Admiral Bobby Inman struggled for over a year to bring first-rate talent to the facilities in Austin, Texas. Then a long silence ensued.

That silence has only recently been broken. MCC is now drawing cautious sketches of the technology it has transferred to its "shareholders" (the 21 companies that invest in and preside over MCC). Despite the slump in the computer industry, the consortium's membership has now more than doubled, and includes such giants as Honeywell, 3M and Control Data. MCC's budget has reached \$65 million a year and its staff is near capacity with 440 employees.

Seven programmes are under way, focusing on semiconductor packaging, software technology, computer-aided design and advanced computer architecture. Each company pays \$500,000 for a share in MCC, and then foots a portion of the bill for each programme in which it participates.

The shareholders also contribute staff — approximately 35 per cent of MCC's employees are "on loan" from shareholder companies. One representative from each company joins the board of directors and one the technical advisory board. Among developments that have already reached shareholders are a model for building expert systems, an editor to aid graphic design of semiconductor circuitry and an improved process for tape-automated bonding which packs silicon chips tightly together on circuit boards and enables denser interconnections to be made.

The new technology is important, but what excites industry is whether the research structure of MCC can provide a new model. Inman says he and his managerial crew have vowed to minimize bureaucracy — "an everyday battle" — and to nurture the type of environment that encourages creativity. The three-level management hierarchy permits autonomy and quick decision-making within research ranks.

And MCC scientists are freed from the vagaries of market forces. The ability to provide stable funding and support is just part of the "psychic income" that keeps MCC laboratories alive, says vice-president and chief scientist John Pinkston.

MCC has invested heavily in its staff: \$20,000 to \$70,000 each for private computer workstations that in typical laboratories would be shared by up to four people. Although the aim is pre-competitive, collaborative research, the opportunities for collaboration are not unlimited. Because all programme information is considered proprietary unless declared otherwise, informal exchange between different programmes is discouraged.

Similarly, MCC scientists must bear in mind the expectations of their shareholders. Although they acknowledge that its goals are long-term, the shareholder companies are anxious to see MCC pay off. But shareholders have so far been reluctant to provide the brightest of their own staff for MCC.

Inman cites technology transfer as the most pressing problem confronting MCC. "We are delivering technology at a level of complexity that not too many of these companies are used to dealing with," he claims.

To enhance communication, quarterly updates and semi-annual research reviews bring shareholder and MCC scientists together regularly. Project teams throw "coming out parties" for newly-assembled technology packages, attended by all participating shareholders. Seminars and technical meetings take place frequently throughout the year. 3M and DEC have even moved their research and development headquarters to Austin.

Will MCC survive? Several industry consortia that have sprung up in MCC's wake pose no challenge, but the multi-billion-dollar budgets of the research and development teams at IBM and AT&T Bell Laboratories dwarf MCC's allotment. And the extent to which companies will capitalize upon MCC's legacy remains uncertain. Unequal gains could force out some shareholders and cause rancour among those that remain.

Finally, the diversity of interests represented at MCC may curtail its expansion into key programmes that would help fill the gap created by such a shake-out. One example is Inman's drive for a manufacturing technology programme, which "raised a lot of welts" at a technical meeting last spring, according to one person who attended. The proposed programme is both too specific to draw broad support among shareholders, and too close to applied science to prevent competitive urges.

But the adventurous spirit that created MCC may sustain it through the transitions ahead. With its unique position in industry research, the consortium has a rare opportunity to test alternatives.

Karen Wright

AIDS

US wins round in patent row

Washington

A US CLAIMS Court last week dismissed a suit brought by the Institut Pasteur in Paris against the United States — representing its National Cancer Institute (NCI) — over the development of a blood test for antibodies to the virus causing acquired immune deficiency syndrome (AIDS). Lawyers for the French research institute expressed confidence that the claims court decision would be reversed on appeal. But the legal issues raised by the Pasteur lawsuit may have a chilling effect on all international exchange of research materials.

Pasteur's case is based on a form signed by Mikulas Popovic, a colleague of Robert Gallo at NCI, when he accepted samples of a virus being studied at Pasteur by Luc Montagnier.

On 15 September 1983, Montagnier presented data at a Cold Spring Harbor symposium attended by Gallo and Popovic about a virus he called LAV. At the time, both Gallo and Montagnier were looking at a retrovirus as a possible causative agent for AIDS. On 23 September, one week after the Cold Spring Harbor meeting, Popovic signed a receipt accepting two isolates of LAV (Mkt-1B and JBB LAV) as well as anti-interferon sheep serum. The receipt stipulated that these materials would be used only for biological, immunological and nucleic acid studies, and would not be used for "industrial purposes without the prior written consent of the Director of the Pasteur Institute". NCI and Gallo have maintained that they adhered to those conditions. But Pasteur's lawsuit claims that Gallo and his associates used the LAV isolates in research that has since led to development of a commercial diagnostic test kit.

The claims court refused to decide the issue of whether the LAV isolates were used improperly. Instead, Judge James Merow based his ruling on whether or not the document Popovic signed constituted a contract with the United States. NCI argued that neither Popovic nor Gallo had authority to commit the United States to a binding contract. The court decided that even if the receipt did constitute a contract, the French institute had not followed appropriate procedures for settling contract disputes — in this case first submitting a claim to the Department of Health and Human Services — and dismissed the case.

James Swire, a lawyer for the Institut Pasteur, described the ruling as "road-block, nothing more". Swire says it is a "sad day for international scientific research" when NCI disavows agreements of its principal investigators.

Reaction from the Department of

Health and Human Services was low key. A spokesman said the department was "pleased" by the court's ruling, and expressed hope that it would "provide impetus to resolving this matter". The claims court decision has no direct effect on a separate proceedings being conducted by the US Patent Office on a disputed patent claim for the AIDS antibody blood test (see *Nature* 320, 96; 1986). The Patent Office has awarded a patent to Gallo for his version of the blood test, but that patent could revert to Montagnier and the Institut Pasteur pending the outcome of the proceedings.

The claims court decision raises important questions about the legal standing of documents signed by researchers when exchanging research materials. J. Edward Rall, deputy director of the National Institutes of Health (NIH) for intramural re-

search, says collaborative agreements are usually entered into on a person-to-person basis. But Rall says most researchers at NIH are aware that there are certain documents they cannot sign.

In 1981, NIH associate director Philip Chen drafted a sample response for NIH scientists to use when requested to sign agreements accompanying cell lines. Researchers must reply that they cannot sign agreements containing waivers or indemnification agreements. They can, however, promise not to share the cell lines without permission from the supplying laboratory.

Scientists have not had to worry about the legal aspects of cooperation before. Both Chen and Rall say litigation is making some grow wary. Rall worries that if things grow worse lawyers may "subvert straightforward scientific collaboration" with legal arguments that will not be appreciated. "It may be that no one will sign anything any more", says Chen.

Joseph Palca

Eureka

Soviets feel left out of Europe

THE Eureka conference in London last month revealed a stance contrary to the Helsinki Accords, according to the official Soviet media. The "obvious tendencies" among participants to limit participation in scientific and technological cooperation to West European countries is seen, at best, as an impediment to the close contacts between countries irrespective of political orientation that were the goal of the Final Act, and at the worst as a means of diverting the supposedly peaceful Eureka programmes for military ends.

Much play has been made of an alleged French distinction between Eureka and the US Strategic Defense Initiative (SDI): SDI is a programme which has a civilian application, while Eureka is a civilian programme capable of a military application.

According to the official Soviet newsagency TASS, Western "political and industrial circles" have recently shown increased interest in the "military aspects of Eureka", while Aleksandr Bovin, a Moscow commentator, stressed that such "giants" as Siemens, Philips and Thomson want to take part simultaneously in both Eureka and SDI. There have been indications in the world press, Bovin said, that Bulgaria, Hungary, Czechoslovakia, Yugoslavia and East Germany are interested in working "within the Eureka framework", but the Western countries do not wish to admit countries that do not share their "ideas about the supreme values".

The Soviet stance may be due partly to pique. In their own opinion, as an eminent Soviet arms control expert, General Nikolai Chervov, stressed at the Royal

Institute of International Affairs last week, the Soviets have made a major concession to the West by softening their demands on SDI: instead of requiring that any arms control treaty outlaw all SDI-related research, they will now accept the compromise that such research be confined to the laboratory stage. Apart from some ironic questions about what, in the context of SDI, constitutes a laboratory, there has been virtually no Western response to this concession.

As for the alleged wish of the West to exclude the socialist countries from Eureka, the problem is a far wider one than a single programme. The whole idea of Eureka, as Bovin rightly noted, is to deal with leading edge technologies — lasers, computers, robots and biotechnology — precisely the fields affected by the Co-Com embargo on technology transfer. Furthermore, even if Eureka does include some programmes that would not come under Co-Com, the Socialist countries cited by Bovin have been remarkably quiet about their wish to participate.

The exception in the socialist bloc is East Germany whose leader, Erich Honecker, has publicly expressed some interest in participation, possibly within the framework of the existing intra-German trade arrangements, and, to a lesser extent, Yugoslavia (which technically ranks as a non-aligned country), where an initial interest now seems to have been replaced by a tacit acknowledgement that, in its state of endemic economic crisis, the country simply cannot afford to participate.

Vera Rich

Patent law

Boost for Canadian drug research

Washington

AFTER a long and painful gestation, the Canadian government has finally given birth to proposals for changes in patent law. The amendments, actively sought by the pharmaceutical industry, will restore to patent holders exclusive rights to sales of patented drugs. In exchange, the government will require pharmaceutical companies to double their current research commitments in Canada.

Since 1969, Canada has had a unique law that allows generic drug manufacturers to obtain licences to produce patented drugs at any time. Once approved for sale by the health protection branch of the Department of Health and Welfare, generic manufacturers are in most cases obliged only to pay a 4 per cent royalty to the patent holder. Major, multi-national drug companies argue that this "compulsory licence" system discourages domestic investment and cuts sharply into profits. But for consumers, the benefits are obvious. A government report last year estimated that Canadians are saving \$220 million a year in drug costs under this arrangement.

Knowing that changing the compulsory licence system will provide political hay for the opposition, the government has several times delayed introduction of the amendments. But as the government has a 70-seat majority, the bill should pass into law unscathed. The government hopes to deflect some consumer criticism by establishing a new Drug Prices Review Board that can remove the protection from compulsory licences for companies whose prices rise too high.

Under the new laws, which will be presented to parliament in the fall but will apply from 27 June this year, companies will not be subject to compulsory licences for ten years. An additional ten years' protection from compulsory licences is offered to companies that manufacture a new drug in Canada. In exchange for such protection, the government has set targets for research expenditure by drug companies. Companies must increase research and development outlay to 8 per cent of sales by 1990, and 10 per cent by 1995. Research expenditure is now running at 5 per cent. Failure to comply with these targets could mean loss of protection from compulsory licences.

Other changes will bring Canada's patent laws in line with practices in most industrialized countries. Following the European model, patent applications will be open to public inspection 18 months after submission. Canada also plans to move away from the US system of awarding patents on a "first to invent" basis in favour of the European "first to file" ap-

proach. The government will also change patent terms from 17 to 20 years, and ratify the patent cooperation treaty.

While pharmaceutical industry trade groups have not yet formally endorsed the proposals, at least one pharmaceutical company is enthusiastic about them. Merck & Co. is moving immediately to complete plans for a Cdn\$20 million addition to its research complex in Montreal and expects to begin construction next year. By 1990, Merck estimates it will spend an additional \$100 million on research and development as well as creating 50 new scientific posts in Canada.

Space vehicles

A short flight for Japan's shuttle

Tokyo

THIS is not the year for space flight. First it was the US Challenger, then Europe's Ariane, now Japan's shuttle has crashed to Earth seconds into its maiden flight. The craft's loss, however, is not a major setback — the shuttle was only a two-metre plastic model.

Last month, scientists from Japan's Institute of Space and Astronautical Sciences (ISAS), the space institute that earlier this year sent two probes to Halley's comet, test-flew a model of an orbital



re-entry vehicle. The white twin-finned shuttle was released at 1,000 m from a helicopter flying at 200 km per h 5 km off the Akita coast of Japan.

Equipped with a three-axis accelerometer, an angular velocity meter, a geomagnetic attitude sensor, an air-pressure sensor and a microprocessor, the model was supposed to be controlled from the ground and to relay information on altitude, attitude and velocity. But no sooner

While victories are being won in Canada, the pharmaceutical industry has not been so fortunate in its dealings with the United States. For years, US drug companies have been seeking a change in the law to allow them to manufacture and export drugs that are not approved for sale in the United States.

Earlier this year an industry-approved bill reached the floor of the Senate, and the House was poised to consider similar legislation. But at the last moment, opponents added amendments so unpalatable to the industry that it withdrew its support for the legislation. The amendments would have placed tighter controls on export of infant formulas, and required additional notification procedures for exports of certain drugs.

Joseph Palca

was the shuttle set free than it looped the loop, turned into a dive and 35 seconds later plunged headlong into the sea.

A second test a few days later with a back-up model was more successful. After release, the shuttle remained airborne and glided steadily for 4.2 km while descending from 800 to 500 m. But it then veered left, failed to respond to correction signals from the ground, and 48 seconds later dived into the sea.

The ISAS team nevertheless considers the experiments a partial success. The telemetry system performed flawlessly and relayed data to Earth which show that the on-board roll stabilizer and yaw damper struggled valiantly to keep the shuttle on a steady course. But with no more models and a limited annual budget of only 100 million yen (£400,000), no further tests are planned this year.

The ISAS Working Group on Winged Vehicles, set up in 1982, is headed by Professor Makoto Nagatomo and has been carrying out computer simulation studies, wind tunnel tests and the present ballistics tests on Japan's shuttle of the future. Once a working model is made, the group hopes to attach a bigger version to a single-stage S-520 sounding rocket — but that would require an increase of two orders of magnitude in its budget, and funds are not yet forthcoming from the Ministry of Education, Science and Culture.

Beyond that, Nagatomo has outlined plans to develop a Highly Manoeuvrable Experimental Space (HIMES) Vehicle by incorporating a cryogenic engine into the shuttle. The HIMES vehicle could then be used as a self-contained re-usable sounding rocket or even as an orbital re-entry vehicle. But ISAS has no ambitions to build a manned shuttle — that is the preserve of the National Aerospace Laboratory and the National Space Development Agency.

David Swinbanks

West Germany

Terror against technology

Hamburg

KARL HEINZ Beckurts, a renowned physicist and one of the most important research managers in West Germany, was killed by terrorists of the Rote Armee Fraktion (RAF), in Munich on his way to work on 7 July. He and his driver, Eckhard Groppler, were victims of a 10-kg bomb, which was ignited electronically as they passed it in their car on their way from Beckurts' home in the Munich commuter village Strasslack to his Siemens office, where he worked as head of the department of research and technics, employing 36,000 workers.

Beckurts, who was born on 16 May 1930 in Rheydt, studied physics in Göttingen, first at the university and then at the Max-Planck-Institut for physics, where he finished his doctorate in 1956. Two years later he became head of the department for neutron physics and reactor technics at the Kernforschungszentrum in Karlsruhe. He moved to the Kernforschungszentrum in Jülich in 1970, and during his ten-year stay was responsible for its reorganization, introducing institutes for biotechnology and contact surface research. He then went to Siemens and became head of the research department.

Apart from his work there and his participation in the Grossforschungseinrichtungen, he taught in Karlsruhe and as an honorary professor in Bonn and Heidelberg. He is said to have been responsible for Siemens' leap into high-technology research.

Beckurts was also a keen supporter of the nuclear programme and it may be this that made him a target for the terrorists. He had fought for his conviction that, even after the Chernobyl disaster, a stable energy supply for West Germany must include nuclear power in the foreseeable future. Beckurts was also head of the Arbeitsgruppe Kernenergie of the Bund Deutscher Industrieller.

It is evident that in planning the killing the terrorists took into account his support for nuclear power as well as the new political situation in West Germany. There is now not only argument over quitting the nuclear programme, but also clashes between demonstrators and police in Brokdorf and Wackersdorf which prove the potential for violence on both sides and the readiness of a few to carry out criminal actions. RAF apparently seeks support from those who have already decided to abandon legal forms of action. A letter found near the site of the bombing "accused" Beckurts of having supported, apart from the nuclear programme, the US Strategic Defense Initiative (SDI) and the European research programme Eureka, described as programmes to fur-

ther the "strategic reorganization of research and production". It is no surprise that a man like Beckurts was involved in consultation on Eureka and SDI. He had, however, committed himself to SDI, and, in spite of Siemens' participation, was not at all enthusiastic about Eureka.

The assassination was apparently planned long in advance. Beckurts' name had been found on a list discovered in an apartment used by terrorists after the killing of Ernst Zimmermann, head of the

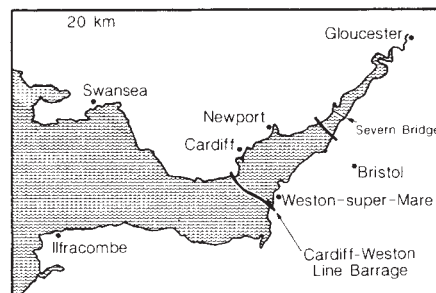
Motor- und Turbinen-Union, last year. Friedrich Zimmermann, Minister of the Interior, told the press that the Siemens manager had been warned. Safety measures will not, however, be tightened up. According to Zimmermann, there is no absolute protection against a remote-controlled bomb: not a very promising outlook for other endangered scientists. But laws regulating the right to demonstrate and the police law may well be tightened up. Political commentators say that this might have been a tactical aim of RAF, because more severe clashes could give them wider support or even new members from whom to recruit.

Jürgen Neffe

Energy

Power from the Severn waters

THE world's largest tidal power-generation project came a step closer to realization last week with the decision of the UK Secretary of State for Energy to back a £5.5 million advanced feasibility programme. The decision was triggered by a new report showing that a 16-km barrage



across the Severn Estuary could be used to generate 14.4 TWh of electricity a year, six per cent of the nation's demand.

The idea of tapping the awesome energy of the Severn's tides — their 30 ft range puts them among the world's highest — has been around for a long time. In 1981, a committee under the chairmanship of Sir Hermann Bondi reported that a barrage was technically feasible. That provided enough encouragement for a more detailed study to be carried out by the Severn Tidal Power Group, a group of six major construction companies, with backing from the government. The report published last week says that the barrage, containing 192 turbine generators, could be built by the year 2000: six years of pre-construction development and seven years of construction, providing 44,000 jobs, would be needed. A further programme will now look in more detail at costs, performance and regional and environmental issues, as well as investigating a smaller tidal barrage scheme for the estuary of the River Mersey.

Two major issues have dominated discussion of the barrage scheme: the cost of the electricity it will generate in comparison with conventional power stations and

the effect it will have on the environment, particularly on the huge area of water behind the barrage.

Serious environmental problems are not foreseen by the report. Water will still flow into the Severn Estuary and it is believed that there would be little change in the salinity of the water or in the rate at which pollutants would move out to sea. But the tidal range behind the barrier would fall by half: although it would still be well within the range found elsewhere in Britain, there is no doubt there would be changes, which are not yet predictable, in the estuary's ecology. Ships using the Severn would be largely unaffected as locks will be provided in the barrage.

Calculating the costs of generated electricity has been more difficult as it depends on making projections into the next century. Electricity would be generated as the tide ebbed and rushed through the line of turbines. Pumping will add more energy: around high tide, when the level of the water either side of the barrier is the same, water can be pumped over the barrier. Later, when the tide has fallen, the extra water can be allowed to run out to give a net energy gain. The gain can be further increased if cheap rate electricity can be used for pumping and the generated electricity returned to the grid at peak rate times. One problem, of course, is that the time when electricity is generated is determined not by demand but by the tides. That, says the Severn Tidal Power Group, is not necessarily a problem: coal-fired stations can be kept on spinning reserve and brought up to full power as the tidal generators slow down.

Taking into account the need for this back-up, the tidally generated electricity seems likely to be marginally more expensive than nuclear-generated electricity but cheaper than coal-generated electricity. As it is unlikely that a large number of new nuclear power stations will be built this century, there should be room for tide power.

Alun Anderson

US education

Liberal colleges ask for more

Washington

EMPHASIZING their "significant and increasingly important contribution to America's scientific manpower pool", a group of fifty leading US liberal arts colleges say they will need to spend an additional \$1,000 million in the next decade to continue providing undergraduate training for future scientists. That conclusion appears in a preliminary draft report writ-

years). The age distribution is skewed, with relatively few faculty younger than 30 or older than 55.

The 1986 report identifies the investments needed to maintain a strong position. Assuming a 1.1 per cent growth in faculty positions will be needed in the next decade, paying for these positions will cost an estimated \$369 million. To continue to attract a strong faculty, \$532 million will be needed to support faculty research, and an additional \$150 million for new laboratories and classrooms.

Appropriate ways of increasing investment in undergraduate science education have been a hot topic at the National

Science Foundation (NSF) recently. A National Science Board task committee chaired by Homer Neal (see *Nature* 320, 479; 1986) called on NSF to increase its support for undergraduate education to \$100 million annually. But NSF must still decide how to implement the Neal report recommendations, and must then win approval for new expenditures from the White House Office of Management and Budget and the Congress.

Carrier believes that the liberal arts colleges would fare well in competition for new NSF grants, but he expects a mix of federal, corporate and foundation support will be necessary to achieve financial goals. A third report to be produced by next summer will assess the issue of where the money should come from.

Joseph Palca

Earned doctorates in empirical and life sciences

Institution	Percentage
Cal Tech	37.5
Harvey Mudd*	37.2
MIT	18.6
Reed*	12.8
UC San Diego	10.8
Swarthmore*	8.2
Haverford*	7.5
Carnegie Mellon	6.4
Wabash*	6.4
Chicago	6.2

Percentage of graduates earning doctoral degrees comparing 50 liberal arts colleges with 20 top rated universities. Starred institutions are among 50 liberal arts colleges.

ten by Sam Carrier and David Davis-Van Atta of Oberlin College, presented last month at a conference on the future of science at liberal arts colleges.

The report is the second in a series assessing the role of small colleges as a training ground for scientists. The 1985 report showed that liberal arts colleges were relatively immune to a nationwide decline in the numbers of new students expressing an interest in majoring in a science discipline. Nearly 30 per cent of entering freshmen at the 50 liberal arts colleges included in the report said they were planning to major in science, compared with a national average of only 5 per cent. Women continue to represent a rising percentage of baccalaureate degrees in the sciences. More than 40 per cent of all undergraduate degrees for women in the 50 colleges were in the sciences.

The report attributes a large part of this strong showing to the close link between teaching and faculty research on the small campuses. Measures of academic distinction show graduates of the liberal arts colleges fare well when compared with graduates from larger universities. Carrier says the importance of liberal arts colleges is the quality of the students they produce as much as the quantity.

A problem for the smaller colleges is a stratification of faculty. At present their science faculties are overwhelmingly male (89 per cent), nearly all white (96 per cent), mostly in tenured positions (69 per cent) and relatively old (average age 44

Chernobyl

Inquest continues on nuclear power

THE Soviet Union is to introduce structural improvements into the design of its nuclear power stations, according to Viktor Sidorenko, deputy chairman of the USSR State Committee for the Supervision of Safe Working Practices in the Nuclear Power Industry. But the type of improvement will be determined only after a detailed study has been made of the causes of the Chernobyl disaster.

In a statement released through the TASS news agency last week, Sidorenko said that his committee had been repeatedly asked, during the past few weeks, whether every reactor should have a dome-type containment building, which would prevent the emission of radioactive material into the external environment. Before Chernobyl, the Soviet stance had been that such containment buildings were unnecessary, and were merely a way for capitalist constructors to increase their profits. Sidorenko, however, took a slightly different line: that containment buildings are simply an alternative approach to safety, and that the Soviet system of modular "strong boxes", as used at Chernobyl, is a valid equivalent. It would be difficult to say at this stage, Sidorenko stated, whether a containment building would have reduced the scale of the Chernobyl emission.

Whether the two approaches are truly equivalent is a moot point. As a recent open letter to the International Atomic Energy Agency (IAEA) from an unofficial team of Polish scientists stressed, Soviet reactors have only two levels of containment, not as in the West, three. The more cautious purchasers of Soviet VVER reactors — Finns, Hungarians and Poles — have therefore insisted on a Western-style containment building being incorporated into the design.

Certainly the "modular strong-box"

approach failed to work for the RBMK reactor at Chernobyl, although, as Boris Semenov, deputy chairman of the USSR State Committee for Nuclear Energy, said in an interview with the Moscow *Literaturnaya Gazeta* in June, the design was intended to be sufficient to stop the leakage of radioactive material from the worst possible accident that could be foreseen: an instantaneous transverse rupture of the pressure header of the main circulation pumps.

One of the tasks of the government commission set up immediately after the accident is to determine and report on its cause or causes. Its other, and even more pressing duty, is to "eliminate the consequences" of the disaster. Since this is proving an extremely complex task, the report, which since the end of May has been promised "shortly", is not yet complete. (The arduous nature of the twofold task has, so far, worn out at least four, and possibly five, commission chairmen, all deputy premiers of the USSR.) Even without the official findings, and in spite of reports of faulty construction at Chernobyl, which appeared in the Ukrainian press before the accident, at least one Soviet official is quite sure that neither the design nor the construction of the Chernobyl plant was to blame.

Two days after Sidorenko's statement to TASS, Aleksei Makhukhin, First Deputy Minister of Power and Electrification, told the other official Soviet news agency, Novosti, that the accident seems to have been due to "a coincidence of several highly improbable and hence unforeseeable failures". Nevertheless, he admitted, although neither design nor construction was at fault, additional measures would be introduced to ensure greater safety of nuclear power engineering.

Vera Rich

A threat to medical progress

SIR—There have in recent years been several examples of a trend that has serious scientific and public importance, and in particular is liable to inhibit the development of medicine.

Any medical or surgical intervention carries, as is accepted, some degree of risk. Before devices or drugs are cleared for public use, they are subjected to close scrutiny and extended trials, and are released only if the risks are found to be acceptably small in relation to the much more substantial benefits. For example, it is accepted that the use of an intra-uterine device involves some slight increase in the risk of pelvic infection, but substantially less than the risks resulting from pregnancy, and it is equally accepted that all drugs carry some risk of undesired side-effects.

However careful and extensive the trials may be, the release of a device or drug to the public provides data on a very much larger scale; this relationship is inevitable and unavoidable. In due course, further risks may come to light, usually those which are so small that they could not have been statistically significant in the original trials.

What is then liable to happen is that the manufacturer is subjected to lawsuits by people who have nothing to lose, and in particular clients of lawyers in those countries that permit contingency fees. These cases are tried and assessed by judges and juries who are not in general scientifically trained or knowledgeable, and who are unlikely to understand statistics or statistical causality. It is then probable that one or more of these cases will succeed, and it can afterwards be said that "it has been established by a court" that the substance or device is harmful.

The manufacturer is then on the horns of a dilemma. If it does not fight the lawsuits then it loses immediately. If it does fight them it may be put to enormous costs even if it wins, and by resisting the claims it is almost certain to attract hostile publicity, in particular media attention in which it is assumed and reiterated, without evidence, that what has been supplied is unquestionably harmful. The problem is so serious that once anything is even alleged against a supplier of a drug or a medical device, there may be little option than for the supplier to go into liquidation.

If this trend continues, it may well become commercially impossible for companies to develop anything new in medicine, and this would have serious public implications.

Of course some drug companies and other large firms have not been blameless, just as others have displayed a commendably responsible attitude. The problem is that the combination of contingency-fee

lawsuits and media emotionalism is almost entirely indiscriminating between those that have behaved well and those that have behaved badly. False sympathies are aroused by the media representing the problem as the little man or woman up against the commercial giant, but it needs to be remembered that it is the large company, rather than the retailer or physician, that is sued because it is only against a large corporation that large damages can be won, amounting in some recent cases to as much as £1 million.

Every sympathy and help is indeed due to someone damaged, in any circumstances, by medical intervention, but indiscriminate treasure-hunt suits do not fulfil this function and they confer a disbenefit on everybody who might need medical help in the future. The courts in this country have generally taken a responsible stand (and been castigated for it) by suppressing reports relating to a case *sub judice*; and when the report is eventually released, it may prove to be not evidence but only hearsay and unsupported assertion. Further legislative adjustment does however seem to be necessary, above all to strengthen discrimination in the assessment of alleged blame.

P.B. FELLGETT

University of Reading,
Department of Cybernetics,
3 Earley Gate,
Whiteknights,
Reading RG6 2AL, UK

Monkeyed about

SIR—Tim Beardsley's News item "Monkey business: Bolivia asks for animals back" (*Nature* 319, 610; 1986) mentions squirrel monkeys and owl monkeys from Bolivia. The scientific names for both species are misspelled. For squirrel monkeys the correct spelling would have been *Saimiri sciureus boliviensis*, not "*Saimiri sciourenous boliviensis*". Further, the appropriate nomenclature for Bolivian squirrel monkeys would have been *Saimiri boliviensis boliviensis*¹.

For owl monkeys, the correct spelling would have been *Aotus trivirgatus*, not "*Actus trivirgatusson*". Further, the correct nomenclature would have been *Aotus azarae*. According to Hershkovitz², virtually all the owl monkeys in Bolivia should belong to one of two subspecies of *Aotus azarae*. *Aotus* exhibits substantial between-population chromosomal variation, and it is clear that it should not be regarded as a monotypic species (as has often been the case in the biomedical research community).

There is a clear need for primates and other animals used in biomedical and behavioural research to be accurately identi-

fied. For imported primates, specific sites of origin should be included in import documents and animal records. Too often scientists do not correctly identify their animal subjects, even in reports published in the best journals. Appropriate identification of primates should be a minimum requirement for their use and importation.

I urge all scientists who use primates to make a special effort to identify the animals they use, including geographic origin, and correctly to report this information in all published materials. For bibliographic assistance, scientists should make use of the services of the Primate Information Center at University of Washington.

J. ERWIN

American Journal of Primatology,
PO Box 65481,
Washington, DC 20035-5481, USA

1. Hershkovitz, P. *Am. J. Primatol.* 6, 257–312 (1984).
2. Hershkovitz, P. *Am. J. Primatol.* 4, 209–243 (1983).

Tim Beardsley replies: The names were copied faithfully from the US Fish and Wildlife Service's import documents; there is no universally recognized taxonomy of squirrel and owl monkeys. □

Soviet computers

SIR—Vera Rich reports skilfully on the planned changes in the administration of Soviet higher education (*Nature* 321, 716; 1986). Quite apart from the fact that one shudders a little whenever the Central Committee (or even someone like Sir Keith Joseph) announces that "we have plans for you", the situation in Soviet science is actually pretty grim for those who have to put up with it (even when they have all the privileges that come with the high status of an "Academician").

For example, at a recent meeting of the Academy of Sciences in Moscow, a very eminent scientist complained in public that not only was he short of computing power, which put him two or three years behind his American opposite number, but he gained a distinct impression that there were influential people around in Soviet science who did not think computers were essential. As a result he had to resort to "unusual ways" of acquiring his computer facilities (one imagines, either by smuggling them out of Western countries, or using the telephone to communicate with a computer in the United States).

The fact is that Soviet science is run by an establishment that makes our own University Grants Committee and Science and Engineering Research Council look like a collective of enterprising young men.

S. CHOMET

Department of Physics,
King's College London (KQC),
University of London,
Strand, London WC2R 2LS, UK

Test ban made more respectable

A new study suggests that even the smallest nuclear explosions cannot be successfully hidden in holes beneath the ground, which makes a test-ban treaty a better proposition.

WHETHER or not there is eventually a comprehensive ban on nuclear tests, it will never be possible for politicians and others to claim that the academic community has failed to give the matter the attention it deserves. So much should be clear from the long account which occupies the first 71 pages of the current issue of *Reviews of Geophysics*, published by the American Geophysical Union, of the argument by which J.F. Evernden and two colleagues have been able to show that even attempts to hide underground explosions of nuclear weapons by conducting them in underground cavities cannot be successful (24, 143; 1986).

The question whether nuclear explosions might be concealed from the prying seismometers of invigilators first became a public issue in the 1950s, in the years preceding the abortive negotiations of a fully comprehensive test ban. In the event, the three governments in negotiation of the question, the Soviet Union, the United Kingdom and the United States, settled instead for a ban on all but underground tests, partly at least because of the belief that sizeable underground explosions might be effectively concealed by "decoupling" them from the surrounding medium. People were set to work ingeniously devising paper schemes for building spherical cavities in salt mines by solution mining and other methods. There was talk of muffling the seismic noise generated by explosions by factors of 50 or even 100. In the more recently aborted negotiation of a comprehensive test ban, in 1979, the three negotiating powers seem to have agreed to set aside the difficulty of concealment by decoupling, but the question has more recently been given a new lease of life and would no doubt be a stumbling-block again if serious negotiations on a comprehensive test ban were under way.

Not surprisingly, the question has an intrinsic as well as forensic interest. What really happens to the seismic signal of an underground explosion carried out in a cavity of some kind? Evernden is at the US Geological Survey at Menlo Park in California; the fellow authors of his latest contributions to the literature of the test ban are C.B. Archambeau (University of Colorado) and E. Cranswick (US Geological Survey), both at Boulder, Colorado.

The conclusion is heartening. According to Evernden *et al.*, it should now be possible to detect underground explosions in, say, the Soviet Union yielding explo-

sive energy equivalent to 1 kilotonne of conventional explosive even when these are fully decoupled from the surrounding rock. This, the authors say, could be accomplished by means of 15 monitoring stations outside the Soviet Union and by a network of 25 automatic seismic stations within the territory of the Soviet Union if they were so disposed as to make the most efficient use of the data gathered. The conclusion is heartening because, in the most recently aborted set of negotiations on the subject, the two major powers had agreed that each would provide house room for exactly 10 such stations.

How has this cheering result been obtained? Those who read the article by Evernden *et al.*, as Evernden's earlier work on the same subject, will be struck by the way the argument is founded uncontentiously on empirical data. The starting point, for example, is a widely accepted formula (due to J.A. Sharpe and first published in 1942) describing the displacement field caused by an explosion in a supposedly homogeneous elastic material as a function of distance and of the frequency of the signal.

The physics of the process is well understood but none the less interesting for that reason. There is no doubt that an explosion in a cavity can in principle muffle the seismic signals from an underground explosion, partly because of the energy absorbed by the compression of the surrounding medium (air) and partly because of the sheer size of the cavity, which may imply that the pressure of the expanding explosive wave may be less, by the time that it reaches the walls of the cavity, than that required to stimulate shock-wave propagation into the rock. And the result of that is that the wave from a decoupled explosion expands outwards, when it reaches the surrounding rock, even less efficiently than will the wave from a well-tamped explosion.

It is good to be reminded by the article of Evernden *et al.* that the US authorities have in their time been obliging in their willingness to allow these predictions to be verified experimentally. Indeed, the authors make good use of the experiment conducted beneath the Nevada desert in the 1960s, when one nuclear explosion was used to make a cavity and another, smaller and later, was used to test the then current understanding of decoupling.

The essence of the argument now published turns on the dependence of dis-

placement at a distance from an explosion of any kind, decoupled or not, on the frequency at which the seismic signals are measured. Simply, the amplitude of the seismic signals is roughly a constant up to a certain value and then appears inversely proportional to the square of the frequency. Moreover, this "corner frequency" is itself a function of the magnitude of the explosion.

The practical consideration is that while the reduction of the seismic amplitude caused by decoupling in the low-frequency range may be very large, amounting to a factor of as much as 100, the disparity in the high-frequency range may be much less, a factor of 10 or so. What Evernden and his associates now propose is that the detection of decoupled explosions will be most efficient in the high-frequency region. Although the magnitude of the seismic displacements is itself reduced by the inverse-square variation of displacement with frequency, modern instruments could cope with that. But the reduced signal strength at higher frequencies explains why it is necessary to increase the number within the Soviet Union from the 10 talked of in 1979 to 25 now.

This latest contribution to the literature of the detection of decoupled explosions will surely not be the last on the subject. Indeed, Evernden and his colleagues raise hares fit for half a dozen investigations. There is, it seems, a possibility that decoupled explosions may be distinguished from well-tamped explosions, raising the possibility that attempted violations of a comprehensive test ban might be detectable. The limits of the sensitivity of detection arise, it turns out, from the confusion caused by wind-driven micro-seismic activity; the network of 25 internal stations is so chosen as to make use of the efficient propagation of high-frequency signals through the shelf structures of mainland Asia. Perhaps the most serious drawback in a system for detecting explosions as sensitively as that now proposed is that quite modest explosions of conventional chemicals would be detectable, raising the awkward possibility that the proposals for on-site inspection of suspicious places, one of the unprecedented provisions of the most recent draft treaty, would too often raise demands for inspection when none was called for. But that is not a serious obstacle to an agreement between governments prepared to take modest risks.

John Maddox

Origin of life

RNA and hot-water springs

from E. G. Nisbet

THE discovery that RNA molecules can extrude introns which can then act as enzymes (refs 1,2; see the recent *News and Views* article³ by Frank Westheimer) makes the notion that life began from RNA very attractive. The most likely site for the inorganic construction of an RNA chain, which would have occurred in the Archaean, is in a hydrothermal system. Only in such a setting would the necessary basic components (CH_4 , NH_3 and phosphates^{4,5}) be freely available. Suitable pH (fluctuating around 8) and temperatures around 40°C are characteristic of hydrothermal systems on land. Furthermore, altered lavas in the zeolite metamorphic facies, which are rich in zeolites, clays and heavy metal sulphides, would provide catalytic surfaces, pores and molecular sieves⁶ in which RNA molecules could be assembled and contained. If the RNA could then replicate in such a setting with the aid of ribozymes and without proteins, the chance of creating life becomes not impossible but merely wildly unlikely.

Models of self-replication

A self-replicating von Neumann machine needs a list of instructions (DNA and RNA in life); a set of organs (proteins) to help in the replication of the list; and a container which protects a local degree of order so that the components can find each other. In some models the container is dispensed with by the assumption that the oceans were a 'soup' of organic molecules. This is most improbable, as the Archaean oceans were probably rapidly processed through lavas in submarine hydrothermal systems at temperatures of several hundred degrees.

In the modern Earth, a volume of sea water equivalent to the entire oceanic volume passes through the mid-ocean ridge hydrothermal systems in about 10^7 years⁸, whereas in the early Archaean this circulation time was probably 10^6 years or less^{9,10}. Unless the production rate of organic molecules in the ocean and atmosphere was high, the soup was probably clear and with chemistry not greatly different from today. Furthermore, NH_3 and CH_4 would not be present in any significant amounts as they would have had very short lifetimes in the atmosphere¹¹. It therefore seems unlikely that life began in the open ocean.

In other models¹², the nucleic acid 'list' is dispensed with and proteins are called on to replicate themselves, but in this case it is difficult to see how the mechanism of replication could have the necessary specificity. The subsequent evolution of

nucleic acids would have to be from information in polyamino acids, which is not according to the central dogma. In an alternative model¹³, the list is based on clay minerals which can replicate themselves and do not need a protective container; but it is not clear how a daughter could detach itself from the parent and assume an independent existence.

In contrast to these interesting but complex models, the discovery that RNA can splice out a length of itself, which can then act as an enzyme^{1,3}, suggests a much simpler path to life. RNA is not only a list, but can also help⁷ to replicate the list. What is needed to complete the model is the third component, a suitable container or environment in which an RNA molecule could be assembled and could detach daughter ribozymes, yet still retain access to them.

A shallow-level hydrothermal system is the most likely setting in which all the necessary components could have been assembled to create a polynucleotide. First, only in hydrothermal systems could significant partial pressures of CH_4 and NH_3 exist in the Archaean Earth.

Second, hydrothermal systems carry large quantities of phosphorus, which has been dissolved out of basalt glass⁵. Typical mid-ocean ridge basalts contain 0.1–0.2 per cent P_2O_5 , virtually all of which is lost to solution on the passage of hydrothermal fluids⁵. In the earliest Archaean the flux released by this route would have been much greater than today, and would have dominated the global supply of phosphorus, especially in the absence of a large sedimentary inventory on the continents.

The third reason for favouring a hydrothermal setting is the availability of inorganic catalysts. Altered basalts and pillow breccia consist of a porous fabric which includes a complex microcrystalline aggregate of zeolites, clay minerals and heavy metal sulphides, as well as albite, carbonates and relict primary phases. Enormous volumes of water flow through such altered rocks, with varying pH, chemical content and temperature, depending on the activity of the driving volcanic heat source and the local geology. Zeolite minerals can act as molecular sieves⁶, allowing the passage of small molecules while retaining large molecules. They also have remarkable and highly specific catalytic properties similar to enzymes¹⁴; the catalytic effects of clay minerals^{15–17} are well known. The abundant pores and cavities in altered basalt would provide an extraordinary assortment of inorganic containers lined with

zeolite molecular sieves, ranging in size from the zeolite cage⁶ (about 11 Å) to cavities several centimetres long.

If a molecule of RNA capable of self-replication formed in this setting it would be in a fabric of altered rock, probably in a pore surrounded by a molecular sieve of zeolite. The pH of some Icelandic hydrothermal systems fluctuates in the range 7.5–8 (in contrast to submarine systems which are acid); such systems also have large regions at temperatures around 40°C; and the concentrations of Mg^{2+} and Cl^- ions in the hydrothermal systems would favour replication of RNA.

The RNA world

In a recent *News and Views* article¹⁸, Walter Gilbert postulated that the first stage of evolution was in an RNA world. In this model world, life consisted of self-replicating RNA molecules and used recombination and mutation to explore new functions and adapt to new niches. Eventually a wide range of enzymic activities would develop by using RNA co-factors. Later, RNA would begin to bind and synthesize proteins which would eventually take over the enzymic role. A hydrothermal system would be ideal for providing the amino-acid components in this process¹⁹. In such a world, replicating systems in pores in altered rock close to the water-rock interface could pass from pore to pore in the altered lava, by chance as hydrothermal systems varied. They would eventually colonize their habitat to the extent of available resources. Any molecule which could replicate more faithfully by isolating itself from the random catalytic effects of zeolites and clays (once so useful, by now a source of error) would be strongly favoured by natural selection¹⁰. A system which learned to surround itself with a bag of lipids would be more successful at reproduction and would be pre-adapted to life in the open ocean. □

1. Zaug, A.J. & Cech, T.R. *Science* **23**, 470 (1986).
2. Guerrier-Takada, C. & Altman, S. *Science* **223**, 285 (1984).
3. Westheimer, F.H. *Nature News and Views* **319**, 534 (1986).
4. Arnorsson, S. & Svavarsson, H. *Geochim. cosmochim. Acta* **46**, 1513 (1982).
5. Staudigel, H. & Hart, S.R. *Geochim. cosmochim. Acta* **47**, 337 (1983).
6. Newsam, J.M. *Science* **231**, 1093 (1986).
7. von Neumann, J. *Theory of self-replicating machines* (University of Illinois Press, Urbana, 1966).
8. van Andel, T.J. H. *Nature* **305**, 178 (1983).
9. Bickle, M.J. *Earth planet. Sci. Lett.* **40**, 301 (1978).
10. Nisbet, E.G. *J. molec. Evol.* **21**, 289 (1985).
11. Abelson, P.H. *Proc. natn. Acad. Sci. U.S.A.* **55**, 1365 (1965).
12. Fox, S.W. & Dose, K. *Molecular Evolution and the Origin of Life* (Dekker, New York, 1977).
13. Cairns-Smith, A.G. *Genetic Takeover* (Cambridge University Press, 1982).
14. Wright, P.A. *et al.* *Nature* **318**, 611 (1985).
15. Bernal, J.D. *The Origin of Life* (Weidenfeld and Nicholson, London, 1967).
16. *Genesis* 2.7.
17. Odom, D.G. *et al.* *J. molec. Evol.* **12**, 365 (1979).
18. Gilbert, W. *Nature News and Views* **319**, 618 (1986).
19. Corliss, J.B. *et al.* *Oceanologica Acta Spec. Pub.* **59** (1980).

E.G. Nisbet is in the Department of Geological Sciences, University of Saskatchewan, Saskatoon, S7N 0W0, Canada.

Cosmology

What makes nearby galactic clusters all move as one?

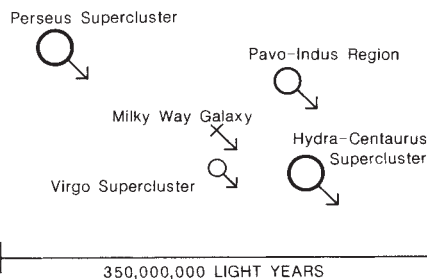
from Joseph Silk

No matter how hard cosmologists try, they have been unable to come up with a simple scenario that simultaneously explains both galaxy formation and the large-scale frothy structure of the galaxy distribution. One class of models, characterized by cold dark matter in the form of weakly interacting elementary particles, does well at explaining small-scale structures, but stumbles at the hurdle of the clustering of galaxy clusters. Another class, characterized by hot dark matter in the form of massive neutrinos or dark baryons such as burnt-out stars or black holes, shows promise of matching the large-scale structure, but at the cost of failing to form galaxies early enough in the Universe. The one point of consensus is the existence of pervasive dark matter, although the lack of any evidence as to its nature has opened the floodgates of speculation. More exotic avenues are being explored, the most promising of which involves cosmic strings, relics from the very early Universe which can act as seeds for the growth of structure on both small and large scales (see Hogan, C. *Nature News and Views* **320**, 572; 1986).

It is premature to tell whether strings will be the cosmic panacea. But no sooner have cosmologists recovered from the momentous discovery of the bubble-like structure of the galaxy distribution on scales of tens of megaparsecs (Mpc) (see my *News and Views* article, *Nature* **320**, 12; 1986), a remarkable new result is being announced. That our Local Group of galaxies is whizzing through space at a speed of about 600 km s^{-1} towards a direction not far from the southern constellation of Centaurus is well known. This motion shows up as an unambiguous dipole anisotropy in the cosmic microwave background (CMB) radiation, seen as a slight heating of the radiation in the direction in which we are moving and as a slight cooling in the reverse direction. Recently it has been reported that our Local Group is not a lone wanderer in intergalactic space, but that galaxies throughout a vast region of about 100 Mpc in extent are companions in its headlong rush. Evidence for this alarmingly coherent large-scale flow has been independently found by two groups.

David Burstein (Arizona State University) and co-workers from six other institutions spanning the globe from Pasadena to Herstmonceux (R. Davies, A. Dressler, S. Faber, D. Lynden-Bell, R. Terlevich & G. Wegner) have presented data on some 400 elliptical galaxies evenly distributed

on the sky. The measured parameters (central velocity dispersions, effective radii and total magnitudes) defined a sufficiently good correlation that enable distances to be inferred once distance-independent parameters (such as velocity dispersion and total magnitude) have been measured. This survey, the results of which were presented at a recent meeting in Hawaii and reported in the *News and Views* article by Richard Bond and Sidney van der Bergh (*Nature* **320**, 489; 1986), shows that there is a bulk motion of about 700 km s^{-1} of essentially all galaxies within $60 h^{-1}$ Mpc of the Local Group ($h = 1$ if the Hubble constant is $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$, or $h = 1/2$ if the Hubble constant is half of this value). The direction of this motion is towards galactic longitude $l = 299^\circ$ and latitude $b = +1^\circ$ and within 20° from the apex of the dipole motion which yields the Local Group motion relative to the CMB.



A traveller at rest with respect to the microwave background radiation looking down on the superclusters would see them moving with about the same velocity in about the same direction. The arrows represent, to scale, the velocities the traveller would see.

(Graphic by David Burstein.)

Another group, working at Imperial College London (Collins, C., Joseph, R. & Robertson, N. *Nature* **320**, 506; 1986) obtained infrared photometry for an all-sky sample of the class of galaxies termed Sc at a mean distance of $50 h^{-1}$ Mpc. Again, by adopting the known correlations between their derived distance-dependent parameter (infrared luminosity) and distance-independent parameters (infrared colour or central velocity dispersion) either absolute distances or Hubble velocities can be inferred once the correlations are calibrated. Comparison of Hubble velocity with the observed recession velocity for about 45 galaxies led Collins and colleagues to infer a bulk streaming velocity of $970 (\pm 300) \text{ km s}^{-1}$ in a direction towards $l = 305^\circ$, $b = 47^\circ$ that, within their large quoted uncertainty, is consistent with Burstein *et al.*'s result.

To many astronomers, this is a case of *déjà vu*. The large-scale motion amounts to a confirmation of the Rubin-Ford effect, a motion of the Local Group relative to a shell of Sc galaxies at $\sim 50 h^{-1}$ Mpc that, following its initial discovery in 1976 by Vera Rubin and Kent Ford, has generally been disbelieved until now. After all, the inferred direction of motion was nearly orthogonal to the apex of the Local Group motion relative to the CMB. This neglect may have been unjustified. Indeed, the first indication of a large-scale bulk motion came with the confirmation by several groups that the Local Group is falling towards the Virgo cluster at $\sim 250 \text{ km s}^{-1}$ and in a direction some 45° away from the apex of its CMB motion. This means that, in the reference frame of the CMB, the entire Virgo Supercluster, including the Local Group at a distance of some $10 h^{-1}$ Mpc from Virgo, is moving at $\sim 400 \text{ km s}^{-1}$ in the direction of Hydra-Centaurus.

The new data on ellipticals tell us that clusters in Hydra-Centaurus in the south, and in Perseus and Pisces in the north, separated by a distance of $\sim 60 h^{-1}$ Mpc, are all sharing a common motion with Virgo, whose amplitude may be as large as 700 km s^{-1} . We are all moving towards a region behind the Hydra-Centaurus clusters. According to Burstein *et al.*, there is an additional component of random motion of $\sim 300 \text{ km s}^{-1}$ for individual clusters superimposed on the bulk flow. Finally, of course, the CMB must represent the ultimate local rest frame. It is reassuring that A. Yahil, D. Walker and M. Rowan-Robinson have confirmed in a study of IRAS galaxies that at a depth of $\sim 200 h^{-1}$ Mpc, the galaxies are indeed at rest with respect to the CMB (*Astrophys. J. Lett.* **301**, L1; 1986). (Similar results were reported last summer at the Princeton dark matter symposium (in the press) by A. Meiksen and M. Davis.)

Reconciling these new data with existing models of large-scale structure will be a difficult task as no model predicts such large velocities. If the interpretation of the data holds up then something must be pulling us; exactly what is not clear. Indeed, it has been argued by N. Vittorio, R. Juszkiewicz and M. Davis (*Nature* in the press) that confirmation of the large-scale bulk motion will invalidate all hot or cold dark matter models that rely on inflation-generated random phase perturbations of a Friedmann cosmology. Less attractive models may work: these include low-density Friedmann cosmologies containing either hot or cold dark matter, together with primordial seeds that have triggered early galaxy formation. The seeds might be cosmic strings, as proposed by T. Kibble, A. Vilenkin and Ya B. Zel'dovich (see Hogan, C. *Nature News and Views* **320**, 572; 1986), or rare objects that have injected sufficient energy via supernova

explosions to explosively amplify the perturbed mass scale up to galactic dimensions, as discussed by J. Ostriker & L. Cowie (*Astrophys. J. Lett.* **243**, 127; 1981) and S. Ikeuchi (*Publ. astr. Soc. Japan* **33**, 211; 1981).

The one source of solace is that the large-scale bulk motion does seem to be a likely consequence of the existence of the Hubble bubbles. These apparent voids, on scales up to $50 h^{-1}$ Mpc, could have been generated by the large-scale flows which would evidently be present in a hot dark matter-dominated universe. Of course, these flows must have maintained considerable coherence to preserve the well-defined bubble surfaces on which the galaxies are found. If this were the case, the bubble interiors should be genuinely devoid of matter.

There is an alternative, needless to say,

offered by advocates of cold dark matter — namely that the voids are illusory, simply reflecting the large-scale inhomogeneity of the luminous matter, concentrated into great clusters and into ridges of galaxies, from the relatively smooth dark matter distribution. Such a situation could arise if only the highest peaks in the primordial fluctuation spectrum managed to form galaxies. But however this biasing arose, one consequence is inevitable: the large-scale flows must turn out to be an artefact of observational error. Astronomers are now rushing to verify the reality of the large-scale flows: their confirmation promises to mark a turning point in cosmology. □

Joseph Silk is in the Astronomy Department at the University of California, Berkeley, California 94720, USA.

Biochemistry

New role for transfer RNA

from Robert Haselkorn

BIOCHEMISTRY is conservative, and as a rule the same biochemical step is carried out in pretty much the same way throughout the kingdoms, microbes and man alike. It is unusual, therefore, that this rule is broken by a basic step in the biosynthesis of haem and chlorophyll. These molecules are both modified forms of a conjugated ring molecule called a tetra pyrrole nucleus, made by the condensation of several molecules of the amino acid δ -amino-levulinic acid. But it turns out that δ -amino-levulinic acid is made in very different ways in the cells of animals and plants. As Schön *et al.* report in this issue (*Nature* **322**, 281; 1986), a glutamyl transfer RNA molecule participates in this pathway in chloroplasts.

Labelled acetate and glycine were shown 40 years ago (Shemin, D. & Rittenberg, D. *J. biol. Chem.* **166**, 621; 1946) to be the exclusive precursors of the carbon atoms of haem in mammals and, later, in photosynthetic bacteria. Subsequent work showed that acetate entered the tricarboxylic acid cycle and the key step in the synthesis of haem is the condensation of the cycle intermediate succinyl CoA, with glycine, to make δ -amino-levulinic acid. This is then transported from the mitochondria to the cytoplasm where subsequent steps in porphyrin biosynthesis occur. When the radioactive tracer experiments were eventually repeated using plant extracts (Beale, S.I., Gough, S.P. & Granick, S. *Proc. natn. Acad. Sci. U.S.A.* **72**, 2719; 1975), it was found that the carbon atoms of δ -amino-levulinic acid were not derived from acetate and glycine but from α -ketoglutarate via glutamate.

During the past decade it was shown

that glutamate is reduced to glutamic semialdehyde and the latter transaminated to produce δ -amino-levulinic acid in chloroplast extracts. Various mechanisms were proposed for the glutamate reduction, some involving phosphorylated intermediates. Participation of RNA in the conversion of glutamate to δ -amino-levulinic acid was first indicated by work at the Carlsberg Laboratory in Copenhagen and at the University of Iowa (Huang, D.-D. *et al. Science* **225**, 1482; 1984). Evidence that the RNA is a glutamyl-tRNA was obtained independently by W.-Y. Wang in Iowa, S.I. Beale at Brown University and the group of D. Söll at Yale University. In this issue, Söll's group reports that the intermediate in barley chloroplasts is one of three isoaccepting glutamyl-tRNAs. This tRNA, which has the unusual modified base 5-methylaminomethyl-2-thiouridine in the first anticodon position, is the only one that functions as a substrate for the reduction to glutamic semialdehyde. Whether this glutamyl-tRNA also functions in protein synthesis is not yet known. The tRNA is encoded in chloroplast DNA.

The two alternative paths to δ -amino-levulinic acid arose, I believe, separately in the heterotrophic and photosynthetic bacteria on the one hand and in cyanobacteria on the other. Heterotrophic bacteria are the evolutionary precursors of mammalian mitochondria while cyanobacteria are the ancestors of chloroplasts. Succinyl CoA and glycine are available raw materials in heterotrophic bacteria, but succinyl CoA is in short supply in most cyanobacteria, which lack the complete tricarboxylic acid cycle. Reduction of an

organic acid to an aldehyde proceeds nicely via an ester intermediate; what better source of glutamate ester than glutamyl-tRNA, already on hand for protein synthesis? To avoid competition for the same intermediate, the glutamyl-tRNA used to make δ -amino-levulinic acid should be modified to prevent its use in protein synthesis, but whether that is actually the case remains to be determined, as noted above.

The failure to observe incorporation of glycine into porphyrins in plants and most green algae suggests that the chloroplast pathway provides δ -amino-levulinic acid for all biosynthesis in plants. What is wrong with plant mitochondria? Where did they come from? Plant mitochondrial DNA is much larger than its mammalian counterpart, replete with direct and inverted repeats that recombine frequently to generate an array of circular molecules. Some plant mitochondrial DNA sequences are homologous to chloroplast DNA; at least one such sequence originated in the chloroplast because it encodes a major chloroplast enzyme. Other plant mitochondrial DNA sequences have nuclear homologues, suggesting a history of rather promiscuous exchange of genetic material between organelles in the evolution of plants. With δ -amino-levulinic acid supplied by chloroplasts, selection for the mitochondrial pathway in plants would be relaxed and the relevant genes could have been lost in the shuffle.

An exception to the general proposal that plants lack the glycine-succinyl CoA condensation pathway is known. The alga *Euglena*, evolutionarily bizarre in other respects, has been shown to contain both pathways (Weinstein, J. & Belae, S.I. *J. biol. Chem.* **258**, 6799; 1983). However, so far the green algae *Chorella* and *Chlamydomonas*, as well as cyanobacteria, are believed to contain only the plant pathway to δ -amino-levulinic acid. A number of the outstanding questions raised by Schön *et al.* could be approached conveniently by the isolation and characterization of mutants, for which *Chlamydomonas* and *Anacystis* appear to be the organisms of choice. Useful in this connection could be a compound called gabaculine (5-amino-1,3-cyclohexadienyl acid) that probably inhibits the transaminase that converts glutamic semialdehyde to δ -amino-levulinic acid (Gardner, G. & Gorton, H.L. *Plant Physiol.* **77**, 540; 1985). A collection of mutants resistant to gabaculine ought to include a few that affect the tRNA and the tRNA ligase as well as the transaminase. Examination of the δ -amino-levulinic acid pathway in fungi would also be rewarding from an evolutionary viewpoint. □

Robert Haselkorn is in the Department of Molecular Genetics and Cell Biology, University of Chicago, Chicago, Illinois 60637, USA.

Cell biology

An unfolding story of protein translocation

from James E. Rothman and Roger D. Kornberg

THE general rule of co-translational protein secretion and membrane insertion is faced with a growing number of exceptions. Many cases are now known in which translation can be completed before membrane translocation. The difficult enough issue of how a nascent polypeptide traverses a membrane and folds up on the other side has, therefore, given way to the more general and seemingly impossible problem of how a fully folded, soluble protein can pass the permeability barrier of a lipid bilayer. Fundamental insight into this problem comes from the ingenious experiment of Eilers and Schatz, described on page 228 of this issue¹, which implies that proteins unfold as they are imported into mitochondria. The existence of enzymes that unfold proteins has been indicated in other contexts; such activities may have broad significance.

Newly made proteins potentially can cross several translocational-competent membranes in the cell. For proteins targeted for import into mitochondria and the endoplasmic reticulum, the specificity is usually determined by the nature of the amino-acid sequence at the extreme amino terminus. Mitochondrial precursors have a leader peptide or signal sequence that is usually charged and hydrophilic. Proteins targeted to the endoplasmic reticulum, on the other hand, have a signal peptide that is mostly hydrophobic in character.

To test whether a protein unfolds as it crosses the mitochondrial membrane, Eilers and Schatz produced a fusion protein consisting of a mitochondrial signal peptide linked to the amino terminus of a well-characterized enzyme that is not normally translocated across any membrane, dihydrofolate reductase. They could then ask whether the translocation of dihydrofolate reductase into mitochondria (measured in a now standard type of cell-free system) is affected by a small molecule (the inhibitor, methotrexate) that binds tightly to the active site of the enzyme. If dihydrofolate reductase crosses the membrane in its folded form, then the binding of methotrexate should be without effect on transport (the methotrexate would simply be transported as well). But if the enzyme must be unfolded in order to cross, then methotrexate should profoundly inhibit uptake, by stabilizing the folded structure of enzyme.

The results of Eilers and Schatz show that import of dihydrofolate reductase (but not other mitochondrial precursors)

is virtually abolished by methotrexate. Moreover, the K_i for inhibition of import is the same as the binding constant of methotrexate to the active site of dihydrofolate reductase. Direct evidence that methotrexate binding stabilizes the compact, folded form of the enzyme comes from proteolytic digestion studies. Only the amino-terminal signal peptide and not the remainder of the fusion protein is accessible to digestion, whereas both parts of the molecule are degraded in the absence of ligand binding.

The bipartite nature of the dihydrofolate reductase fusion protein, with an exposed amino-terminal signal peptide and compact core, fits nicely with the finding by Schleyer and Neupert² that transport into mitochondria is at least a two-step process. First the amino-terminal signal and some additional sequence traverses the membrane; then the rest of the chain follows. Schleyer and Neupert trapped protein precursors between steps, in the act of translocation, by using low temperatures, both *in vitro* and *in vivo*. In the two cases examined (the beta subunit of F₁ATPase and cytochrome *c*), the polypeptide chain is caught with its amino terminus penetrating the matrix space (shown by cleavage of its signal peptide) but with much of its mass still on the outside of the mitochondrion where it is susceptible to external proteolytic attack. Only the first step of import requires energy, derived from the mitochondrial membrane potential, because the remainder of the process occurs when the temperature is raised in the presence of energy poisons. The energy that drives protein unfolding must be introduced in the first step, even though complete unfolding may only occur later.

A similar multi-stage process involving protein unfolding may operate during protein translocation across endoplasmic reticulum and across bacterial plasma membranes. It was recognized several years ago^{3,4} that protein transport across bacterial membranes can occur post-translationally, as in mitochondria. Only recently, however (as discussed in the recent *News and Views* article⁵ by Schatz), has it become apparent that translocation across the endoplasmic reticulum membrane can also occur post-translationally, although in at least some cases the completed chain must still be attached to the ribosome⁶. This suggests that all forms of translocation may have common mechanistic features⁵, possibly including an

unfolding step.

Previously, all the evidence had suggested that transport across endoplasmic reticulum membranes was unique in being coupled to protein synthesis, implying that already folded chains are unable to cross. It now appears that when a precursor is unable to cross the endoplasmic reticulum post-translationally it is the fault of the precursor and not of the translocation machinery. Such a precursor simply does not put its best foot forward, apparently folding in such a way as to obscure its signal sequence. As would be expected, post-translational import of proteins by the endoplasmic reticulum requires energy, although in the form of a nucleoside triphosphate, rather than a membrane potential. One wonders whether energy is also used here for an initial step in which the signal sequence traverses the membrane.

Is the machinery that drives protein unfolding limited to membranes? If such a machine were free in the cytoplasm and operated under the same principles, it could unfold selected sets of proteins with selectivity being based on the appropriate 'signal' on the protein (analogous to the endoplasmic reticulum and mitochondrial signal sequences). For example, if a proteolytic activity were built into an unfolding machine, then selected proteins would be degraded as they were unfolded in the cytoplasm. Proteolysis would be ATP-dependent because of the need for energy in unfolding, and it would be limited to substrate proteins unfolded within the machine. Indeed, the ATP-dependent protease from *Escherichia coli* that is responsible for the degradation of partially denatured proteins seems to couple unfolding to proteolysis in such a fashion.

This remarkable enzyme, described by Goldberg and colleagues⁷⁻⁹, degrades a polypeptide chain into small, acid-soluble peptides, starting from one end and finishing off the chain it is degrading before it starts another. This processive degradation requires sustained ATP hydrolysis which, Goldberg and co-workers suggest, allows rapid translocation of the enzyme along its polypeptide substrate. Were such an enzyme mounted in a membrane (deprived of its protease activity), it would translocate a polypeptide substrate past itself and across the membrane. The ATP-dependent protease from *E. coli* thus seems to represent the first known protein translocator.

The power of unfolding enzymes could also be harnessed for constructive purposes. Such enzymes could even catalyse the fundamental process of protein folding. Although this process is generally viewed as spontaneous, because isolated polypeptides can re-fold efficiently, artificial manoeuvres such as the gradual removal of a denaturing agent are often required. Wrongly folded structures

otherwise form and persist. Misfolded proteins in a cell might be recognized by virtue of the chain they expose and then be actively unfolded to allow the folding process to repeat itself rapidly until a compact structure of minimum energy is achieved. Such catalysis of folding by unfolding enzymes would set a minimum stability for a correctly folded structure.

The situation would be analogous to the action of helix-destabilizing proteins such as *E. coli* recA^{10,11}, and single-strand binding proteins¹² in catalysing the reassociation of DNA single strands. These proteins presumably bind to the sugar-phosphate backbone of DNA and destroy wrongly paired duplexes of limited size that would otherwise persist for some time, allowing the strands to reassociate repeatedly until a stable structure, paired in the correct register along its entire length, is achieved. Proteins that act in a similar fashion to facilitate protein folding might recognize and bind to the backbone of a polypeptide chain. Such proteins would act in concert with other factors, such as an enzyme discovered long ago, which facilitates thiol-disulphide interchange during protein folding until the correct (most stable) arrangement of disulphide bonds is achieved¹³⁻¹⁵.

The exciting possibility emerges that cells have a set of unfolding enzymes. When they span membranes such as those of the mitochondrion or rough endoplasmic reticulum and have the appropriate signal-recognition machinery to provide the specificity for initial entry of

the protein substrate, they catalyse translocation and unfolding. When a protease activity is built into the unfolding enzyme, selected proteins can be degraded as they are unfolded. One can even imagine an unfolding enzyme lacking proteolytic activity that still recognizes incorrectly folded proteins. Such an enzyme would correct the errors by allowing misfolded polypeptide chains a chance to try again. Is it possible that the rapid acquisition of a three-dimensional structure, one of the few processes in cells that seems to occur without the aid of enzymes, may in fact receive help? This unfolding story is far from wrapped up. □

1. Eilers, M. & Schatz, G. *Nature* **322**, 228 (1986).
2. Schleyer, M. & Neupert, W. *Cell*, **43**, 339 (1985).
3. Date, T., Zwizinski, C., Ludmerer, S. & Wickner, W. *Proc. natn. Acad. Sci. U.S.A.* **77**, 827 (1980).
4. Randall, L. *Cell*, **33**, 231 (1983).
5. Schatz, G. *Nature News and Views* **321**, 108 (1986).
6. Perara, E., Rothman, R. & Lingappa, V. *Science* **232**, 348 (1986).
7. Goldberg, A. & Waxman, L. *J. biol. Chem.* **260**, 12029 (1985).
8. Waxman, T. & Goldberg, A. *J. Cell Biochem.* (in the press).
9. McEntee, K., Weinstock, G.M. & Lehman, I.R. *Proc. natn. Acad. Sci. U.S.A.* **76**, 2615 (1979).
10. Bryant, F.R. & Lehman, I.R. *Proc. natn. Acad. Sci. U.S.A.* **82**, 297 (1985).
11. Christiansen, C. & Baldwin, R.L. *J. molec. Biol.* **115**, 441 (1977).
12. Goldberger, R.F., Epstein, C.J., & Anfinsen, C.B. *J. biol. Chem.* **238**, 628 (1963).
13. Venetianer, P. & Straub, F.B. *Biochim. biophys. Acta* **67**, 166 (1963).
14. Lambert, N. & Freedman, R.B. *Biochem. J.* **228**, 635 (1985).

James E. Rothman is in the Department of Biochemistry and Roger D. Kornberg is in the Department of Cell Biology at Stanford University School of Medicine, Stanford, California 94305, U.S.A.

Neurotransmission

Are there two functional classes of glutamate receptors?

from Charles F. Stevens

A RECENT article in *Nature*¹ and two that appear on page 263 and 265 of this issue^{2,3}, help to clear up some mysteries about excitatory synaptic transmission in the brain. Neurones that respond to the excitatory neurotransmitter glutamate generally possess at least two distinct classes of receptors, classified by the artificial agonists that excite them, into NMDA (for *N*-methyl-D-aspartate) and non-NMDA types. Under normal recording conditions the NMDA type has rarely seemed to operate. The two articles in this issue show that the NMDA receptor is, indeed, used and define some conditions that determine its use, whereas the earlier article shows why two distinct classes of neuronal glutamate receptors may be useful.

Acidic amino acids, especially glutamate, are to the brain what acetylcholine is to the neuromuscular junction. That is,

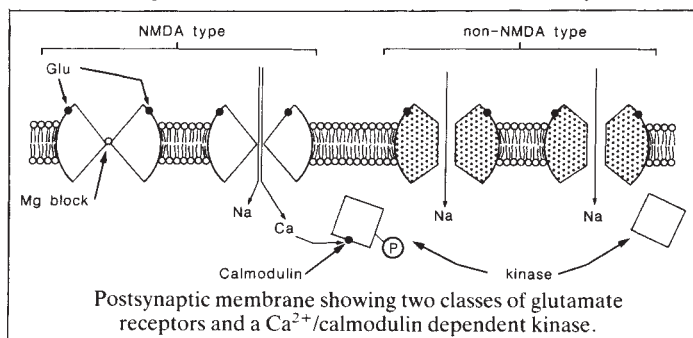
where acetylcholine is used for excitatory synaptic transmission between nerve and muscle, glutamate is used at brain excitatory synapses. It is likely to be that more than half the neurones found in the brain use glutamate as their neurotransmitter.

At least three distinct classes of glutamate receptors can be defined pharmacologically. One class responds to NMDA, whereas the other classes respond to kainate and quisqualate. Some evidence suggests that the kainate and quisqualate receptors may be separate, but these two receptor types

are often classed together and often only two main classes of receptors — NMDA and non-NMDA — are distinguished. Amino-phosphonovalerate (APV) is a specific blocker for the NMDA receptor class but general glutamate receptor blockers that work on all classes are also available. Many neurones have high densities of binding sites for NMDA and are strongly excited when NMDA is applied directly to their surface. Although this excitation is effectively blocked by small doses of APV, this antagonist generally has little effect on excitatory transmission produced by stimulating presynaptic axons. Such excitation, however, is well blocked by broad-spectrum glutamate receptor antagonists. Therefore excitatory synaptic transmission must be normally mediated by the non-NMDA receptor class but what, then, is the function of NMDA class of receptors?

Salt¹ has investigated the contribution of NMDA receptors to synaptic transmission on the thalamus and has found, as is the case of other synapses, that the blocker APV has no effect on the baseline discharge of thalamic neurones and does not alter the response of the thalamic neurones to electrical stimulation of the excitatory pathways that project there. NMDA receptors would thus seem not to be involved in synaptic transmission in the thalamus. However, when natural stimuli are used, for example puffs of air instead of electric shocks, APV almost completely blocks the response of these same thalamic neurones. In other words, NMDA receptors appear to be used when the stimulation is natural but not when afferent pathways are stimulated synchronously with electric shocks.

A clue to how this paradoxical result could arise is provided by the observation of Herron *et al.*³. Although the conditions of their experiments differ from those of Salt, they find, using only electrical stimulation, that the intensity and pattern of stimulation can favour the participation of NMDA receptors in the response of a neurone. Their findings can be explained in several different ways, one of which is that NMDA receptors respond preferentially to low concentrations of glutamate and non-NMDA receptors respond to higher concentrations. By the same explanation, the natural stimulation used by Salt could



perhaps give rise to lower agonist concentrations than would synchronous stimulation, thereby favouring the use of NMDA receptors in one case and non-NMDA receptors in the other. Although further experiments will be needed to sort out the precise mechanisms involved, NMDA receptors are clearly there to be used.

But why are two classes of receptors required? One important difference between the two receptor classes is that the efficacy of synaptic transmission using NMDA receptors (but not the non-NMDA type) depends strongly on the voltage of the postsynaptic cell: if the membrane potential of a neurone is quite negative, NMDA applied directly has no effect whereas NMDA is quite effective in causing excitation (depolarization) if the neurone is already excited (depolarized). This property is, as far as we know, unique to NMDA receptors and arises through an interesting mechanism^{4,5}. NMDA always causes channels to open, but at negative voltages the channels are immediately blocked by physiological concentrations of magnesium ions whereas at more positive voltages the magnesium ions are driven out of the pore and the NMDA channel can excite the neurone by permitting sodium ions to flow in.

The paper by McDermott *et al.* demonstrates another important distinction between NMDA receptor channels and their non-NMDA counterparts: activation of NMDA receptors permits not only sodium but also calcium ions to flow into the cell and to produce significant increases in

the local calcium concentration.

One important consequence of this calcium flux in at least some neurones has been suggested following an observation first made by Saitoh and Schwartz⁶ and recently confirmed and extended by Miller and Kennedy⁷. A kinase regulated by calcium, known to be present in the postsynaptic membrane, appears clustered at many excitatory synapses, so that glutamate action on NMDA-type receptors would produce a calcium influx and consequently activate this kinase. The recent studies reveal that, when sufficiently stimulated by calcium, this kinase phosphorylates itself, and the phosphorylated enzyme then remains active even when calcium concentrations return to resting low levels. Activation of NMDA-type channels then can activate a kinase and switch the metabolic state of the postsynaptic region. Precisely which proteins the kinase phosphorylates, what function they have and whether this effect is a basis for learning and memory as is widely suspected, are questions whose answers should not be long in coming. □

1. MacDermott, A. B. *et al.* *Nature* **321**, 519 (1986).
2. Salt, T. E. *Nature* **322**, 263 (1986).
3. Herron, C. E. *et al.* *Nature* **322**, 265 (1986).
4. Nowak, L. *et al.* *Nature* **307**, 462 (1984).
5. Mayers, M. L., Westbrook, E. L. & Guthrie, P. B. *Nature* **309**, 261 (1984).
6. Saitoh, T. & Schwartz, J. H. *J. Cell Biol.* **100**, 835 (1985).
7. Miller, S. G. & Kennedy, M. B. *Cell* **44**, 861 (1986).

Charles F. Stevens is in the Department of Molecular Neurobiology, Yale University Medical School, New Haven, Connecticut 06510, USA.

Ocean Drilling Program

Palaeoclimatic linkage between high and low latitudes

from the Leg 108 shipboard scientific party*

ONE of the major questions in palaeoceanography concerns the linkage between the polar components of the climate system (ice sheets, sea ice and polar oceans) and the low-latitude ocean-atmosphere components (surface ocean, upwelling regions, wind circulation and land climate). Potential linkages between high and low latitudes have been proposed for time-scales ranging from the early Cenozoic^{1,2} to the late Pleistocene^{3,4}. The central question is whether the two portions of the system are independent or interdependent, and if interdependent, to what degree and through which linkages?

The answer to these questions is important for the current debate on the relative contributions of ice-volume variations and ocean-temperature changes to the $\delta^{18}\text{O}$ record, particularly in low-latitude planktonic foraminifers. It is also important in understanding long-term aridity-

humidity cycles on the African continent, which seem to be driven by orbital insolation variations⁵, but which may also depend on evaporation from equatorial surface waters⁶. In addition, the transfer of carbon from the surface to the deep ocean in productive upwelling regions may provide global climatic feedback via atmospheric CO_2 levels.

Between 21 February and 17 April, Leg 108 of the Ocean Drilling Program (ODP) cored more than 3,800 m of sediment at 12 sites in the eastern equatorial Atlantic (Fig. 1). At the coarse resolution achieved with our shipboard analyses, the initial results suggest a broad-scale late Neogene linkage between the polar and equatorial responses, but further shore-based analyses at the resolution of orbital periods (20,000–100,000 yr) will be necessary to determine the kind of link.

Results from three sites (657, 658 and

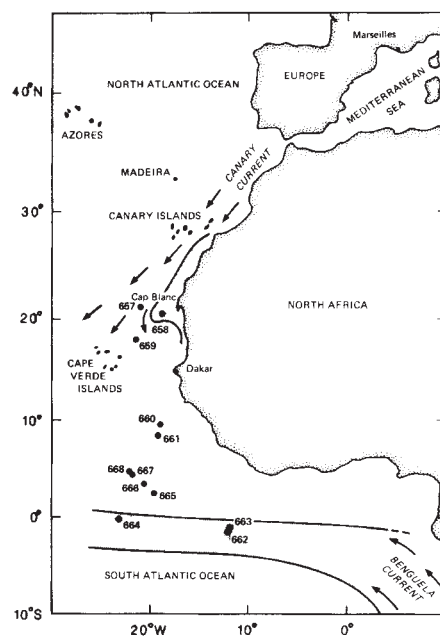


Fig. 1 Sites cored by Leg 108. Arrows, major current systems; dotted areas, regions of strong Plio-Pleistocene upwelling and divergence.

659) in and outside the upwelling system driven by Northern Hemisphere trade winds along the north-west African margin suggest a substantial intensification of coastal upwelling since 3 million years (Myr) ago. A progressive intensification during the past 3–2.5 Myr also emerges from analyses of sediment at three sites along the south equatorial divergence (662, 663 and 664). In both locations, these conclusions are based on increases in the content of opaline silica (diatoms) and organic carbon in layers increasingly rich in terrigenous silt and clay from the late Pliocene to the Pleistocene. These indicators become more abundant in successively deeper Plio-Pleistocene CaCO_3 minima, probably at orbital rhythms (Fig. 2). Our conclusions are further supported by changes in the frequency of cool-indicator nannofossils, planktonic foraminifers and diatoms. In addition to local divergence and upwelling driven by northern and southern trade winds, some of the response in each region may also represent increased equatorwards advection of cool water by stronger eastern boundary currents in each hemisphere (Fig. 1).

The level of resolution of the shipboard

* Co-chief scientists: W. Ruddiman (Lamont-Doherty Geological Observatory), M. Sarnthein (Geologisch-Paläontologisches Institut, Universität Kiel), ODP Staff Scientist, J. Baldauf (Texas A & M Univ.), Also, J. Backman (Univ. Stockholm), J. Bloemendal (Univ. Rhode Island), W. Curry (Woods Hole Oceanographic Inst.), P. Farrimond (Univ. Bristol), J.-C. Faugères (Univ. Bordeaux I), T. Janneck (Lamont-Doherty Geological Observatory), Y. Katsura (Univ. Tsukuba), H. Manivit (Univ. Paris VI), J. Mazzullo (Texas A & M Univ.), J. Mienert (Univ. Kiel), E. Pokras (Lamont-Doherty Geological Observatory), M. Raymo (Lamont-Doherty Geological Observatory), P. Schultheiss (Inst. Oceanographic Sci.), R. Stein (Univ. Giessen), L. Tauxe (Scripps Inst.), J.-P. Valet (CNRS), P. P. Weaver (Inst. Oceanographic Sci.), H. Yasuda (Kochi Univ.).

analytical signals (approximately 50,000 yr) does not allow us to distinguish between whether these trends developed slowly over the past 3 Myr or whether they developed with abrupt step-like changes in amplitude. Regardless, a clear first-order correlation exists between this equatorial trend and that seen in polar climates. Northern Hemisphere ice sheets first attained sufficient size to initiate ice rafting in the subpolar North Atlantic at 2.5 Myr (refs 7,8). Moderate-scale variations in ice-sheet size during the past 2.47 Myr then culminated in the large-amplitude 100,000-yr cycles during the past 650,000 yr (refs 9,10). The changes in equatorial upwelling/divergence show about the same late Pliocene onset and Plio-Pleistocene intensification (Fig. 2).

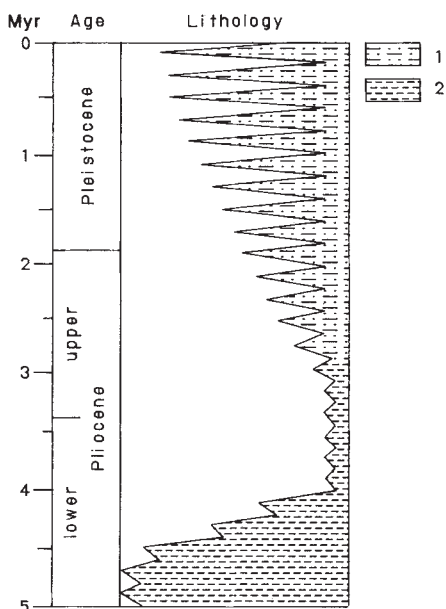


Fig. 2 Schematic log of lithological sequences obtained in and near north-west African margin upwelling area (sites 657, 658 and 659) and equatorial divergence area (sites 662, 663 and 664). CaCO_3 cycles increase in amplitude progressively from 3–2.5 Myr until the present, with non-carbonate lithology consisting mainly of silty clay with opaline (diatom) silica and minor amounts of organic carbon. 1, Silty clay with diatoms; 2, red clay.

In addition to the north polar and equatorial changes, there is also evidence of northwards expansion of the Antarctic Circumpolar Current and the Antarctic polar front at and after 2.7–2.6 Myr (ref. 11). The fact that all three regions experienced nearly simultaneous changes complicates direct cause-and-effect interpretations. Similar first-order synchronicity of responses in polar regions and equatorial oceans has also been noted for the late Miocene^{12–14}.

The exact polar-equatorial linkage thus remains unknown. Eventually, the question will best be answered by orbital-scale analysis of the long Leg 108 sequences using the numerical approaches pioneered by Project SPECMAP¹⁰. At this level of

analysis, the more exact criteria for defining climatic interdependence (analysis of orbital rhythms, phasing and coherence) will narrow the range of explanations.

There are already some suggestions of linkages from orbital-scale studies of conventional piston cores. Late Pleistocene surface-ocean coolings along the north-west African margin appear to vary with the size of Northern Hemisphere ice sheets^{4,15}. In contrast, the equatorial surface waters have a predominant 23,000-yr signal in the late Pleistocene; not only is the dominance of this rhythm different from the distribution of orbital power in the ice sheets, but the equatorial sea-surface temperature response leads ice volume by several thousand years¹⁶. Because the geographical equator lies south of the thermal equator in a region of Southern Hemisphere influence, the equatorial surface ocean may be forced from high southern latitudes on orbital timescales. Thus, the two primary upwelling/divergence regions in the tropical/subtropical eastern Atlantic may be driven from opposite polar regions.

Fine-scale analyses will be aided by new correlation techniques introduced on Leg 108. Previous Deep Sea Drilling Project (DSDP) and ODP palaeoceanographic legs, despite coring two offset holes with the hydraulic piston cores at some sites, have had difficulty obtaining continuous sections; disturbed cores and incomplete recovery of some sections create gaps in the sediment sequence and thus in the climatic record. Only DSDP Leg 94 succeeded in verifying continuity of section in detail (tens of centimetres) at sea by tracking CaCO_3 layering in core photographs and then spot-coring any gaps.

ODP Leg 108 introduced a more sophisticated strategy that should be applicable to future palaeoceanographic legs in regions of large lithological variability. We measured two signals (P -wave velocity and magnetic susceptibility) at intervals of 3 cm or less on unsplit cores, which

allowed correlation of the records at the offset holes to within a few centimetres. As a result, laboratory analyses of the various upwelling/divergence or trade-wind signals are provided with a 'road map' showing precisely which sections to analyse from the offset holes to obtain a complete record of each site. In addition, P -wave and magnetic susceptibility signals contain orbital-scale rhythms and may prove as useful as some of the more conventional palaeoclimatic indicators.

We found evidence of one major Neogene change in palaeo-deep-water circulation: preservation of CaCO_3 was poor before 4.6–3.8 Myr, with red clay below 4,000 and dissolved calcareous sediments above (Fig. 2). After that time, preservation markedly improved, with CaCO_3 preserved to well below 4,000 and less dissolution of the shallower calcareous sections. This change appears to represent the culmination of a cyclical long-term late Neogene trend towards improving CaCO_3 preservation that began in the late Miocene. An accompanying increase in sedimentation rates was more narrowly bracketed at several sites between 4.6 and 3.8 Myr. This change in the CaCO_3 compensation depth is the major equatorial Atlantic deep-water shift of the Plio-Pleistocene. □

1. Shackleton, N.J. & Kennett, J.P. *Init. Rep. DSDP Leg 29*, 743 (1980).
2. Matthews, R.K. & Poore, R.Z. *Geology* **8**, 501 (1980).
3. Cline, R.C. & Hays, J.D. *Mem. geol. Soc. Am.* **145**, 464 (1976).
4. Sarnthein, M. *et al.* in *Geology of the Northwest Africa Continental Margin* (eds von Rad, U. *et al.*) 545 (Springer, Berlin, 1982).
5. Kutzbach, J.E. *Science* **214**, 59 (1981).
6. Flohn, H. *J. Met. Soc. Japan* **60**, 268 (1982).
7. Backman, J. *Stockh. Contr. Geol.* **32**, 115 (1979).
8. Shackleton, N.J. *et al.* *Nature* **307**, 620 (1984).
9. Shackleton, N.J. & Opdyke, N.D. *Quat. Res.* **3**, 39 (1973).
10. Imbrie, J. *et al.* in *Milankovitch and Climate II*, 269 (1984).
11. Ciesielski, P.F. & Weaver, F.M. *Init. Rep. DSDP Leg 71*, 461 (1983).
12. Kennett, J.P. *N.Z. J. Geol. Geophys.* **10**, 1051 (1977).
13. Thunell, R.C. *Mar. Micropal.* **6**, 71 (1981).
14. Denton, G.H. & Armstrong, R.L. *Am. J. Sci.* **267**, 1121 (1969).
15. Thiede, J. *Meteor. Rep.* **C28**, 1 (1977).
16. McIntyre, A., Ruddiman, W. & Karlin, K. *Transact. Am. geophys. Un.* **66**, 931 (1985).

Thyrotrophin-releasing hormone

New applications in the clinic

from Ewan C. Griffiths

As a result of research reported at a recent meeting*, it seems likely that thyrotrophin-releasing hormone (TRH) and its analogues are about to attract interest as novel therapeutic agents in various neurological disorders. TRH, the tripeptide L-pyroglyutamyl-L-histidyl-L-prolineamide (Burgess, R. *et al.* *Nature* **226**, 321; 1970), acts as a hypothalamic hormone in stimulating the release of thyrotrophin (and other hormones) from the anterior pit-

uitary. More than 70 per cent of the total brain content of TRH however, is found outside the hypothalamus and several initial studies suggested that it might be useful clinically as an antidepressant. But none of the analogues produced proved useful, so interest in the hormone declined.

One particular line of research that has re-awakened interest in TRH now is the demonstration that TRH can reverse induced states of shock and spinal injury (J.W. Holaday, Washington DC; A.I. Faden, San Francisco; S. Amir, Rehovot).

*Therapeutic uses of TRH. Santa Barbara, California. 6–8 April 1986.

In endotoxic, anaphylactic, lipoxygenase-induced and leukotriene-induced shock, TRH has proved a remarkable agent in increasing survival and reversing depressed cardiovascular and respiratory function. In spinal cord trauma, this neuropeptide was particularly effective, though not in dog and gerbil models of brain ischaemia. In a rat model of cerebral ischaemia (Latham *et al. Regul. Peptides* **13**, 80; 1985), two TRH analogues given intraventricularly have been dramatically effective: RX 77368 (pGlu-His-[3,3'-dimethyl] ProNH₂) and CG3509 (orotyl-His-ProNH₂).

One area of clinical relevance for TRH is in motor neurone disease or amyotrophic lateral sclerosis. There are improvements in muscle tone and function after intravenous TRH treatment in some, but not all, patients with motor neurone disease and other types of spasticity (W.K. Engel, Los Angeles; B. Brooks, Madison). Other preliminary findings on the improvement of muscular functions in motor neurone disease used intrathecal administration (T. Munsat, Boston). These three studies suffer from short duration of action of TRH which has a half-life of 4 min or less. To overcome this, R. Guiloff (Westminster Hospital, London) has undertaken a clinical trial in patients with RX 77368, an analogue that binds well to TRH receptors and, compared with other analogues, is extremely stable and potent (my own observations). The patients treated with RX 77368 had significant improvements in speech, respiratory function, tongue movements and swallowing and decreased spasticity. There was also a marked increase in force of muscle contraction and in motoneurone activity.

How does TRH and its analogues work in motor neurone disease? Some clues have come from the work of Valerie Askansas (Los Angeles) who finds that motoneurons grown in culture in the presence of TRH show increases in choline acetyltransferase activity and neurite outgrowth, so perhaps the hormone is a neurotrophic agent. Further evidence for a neurotrophic effect has come to light through the reported reversal by TRH of neuronal damage caused by the substance P antagonist spantide (Freedman, J. *et al. Expl Brain Res.* **62**, 175; 1986), and the stimulation of myelin lipid synthesis in chick neural cultures (Kanamoto, Y. *et al. Brain Res.* **371**, 201; 1986). In patients with motor neurone disease, there is a loss of TRH and TRH receptors in the spinal cord (laminae II and IX). TRH is known to excite α -motoneurone activity and may act at several central and peripheral neuronal sites to cause the improvements observed in motor neurone disease.

Further potential applications arise from the variety of central nervous system actions attributed to TRH. In an animal model of Alzheimer's disease (neuronal

damage induced by AF 64A, a nitrogen mustard derivative of choline), another TRH analogue MK771 (L-pyro-2-amino-adipyl-His-thiazolidine-4-carboxamide) was reported by A. Horita to reverse the decrease in choline acetyltransferase activity. Because Alzheimer's disease is believed to involve a cholinergic deficit, perhaps TRH and its analogues could be used to reverse this deficiency. In view of the antiopioid actions of TRH (except in prolactin release and analgesia), it might be useful in treating narcolepsy and, through its analeptic effects, in the reversal of anaesthesia. Similarly, there is no really satisfactory therapy for either stroke or spinal injury and although there may be a time-limit on the effectiveness of the hormone, this is another possible use.

Such a list of previously untreatable

conditions has not gone unnoticed by the pharmaceutical companies, and besides motor neurone disease, clinical trials are being planned or are already under way with TRH and its analogues in spinal injury. Perhaps there is greater scope for the more stable analogues with enhanced duration of action, and choice of the best one from a long list of analogues now available can be assisted by both biological studies and theoretical prediction of conformation (Ward, D.J. *et al. Regul. Peptides* **13**, 73; 1985). After being thought of only as an ineffective antidepressant and a means of testing anterior pituitary function, TRH is making a comeback. □

Ewan C. Griffiths is Lecturer in Physiology at the University of Manchester Medical School, Manchester M13 9PT, UK.

Biological chemistry

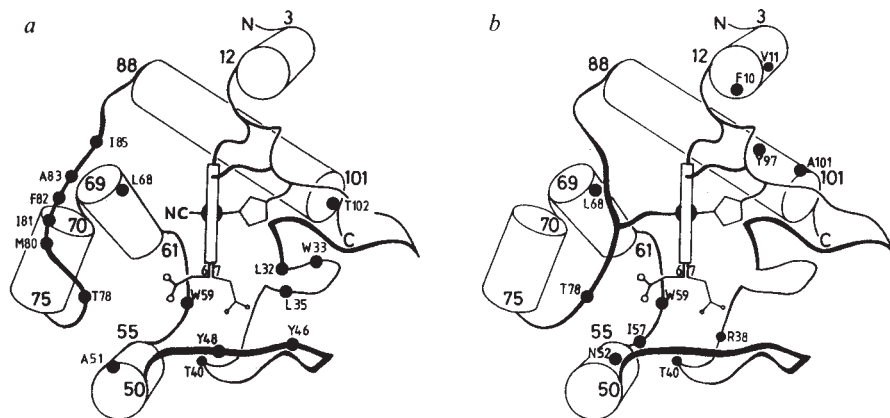
Long-range electron transfer

from R.J.P. Williams and D. Concar

THE basis of many biological processes, including many fundamental steps of energy transduction in membranes, depends on long-range (10–20 Å) electron transfer. The postulation of these electron hops, based on observations of complex biological pathways, was necessary as soon as it was realized some twenty years ago that the metal centres of electron-transfer proteins are embedded in folds some 5–10 Å from the surfaces of the proteins. That such processes were possible was then demonstrated by theoretical and solid-state studies. But the recent results of S. Isied and colleagues, reported elsewhere in this issue (Bechtold, R. *et al. Nature* **322**, 286; 1986), apparently throw

much of our subsequent knowledge of these processes into question.

The concept of long-range electron-hop reactions originally met with scepticism from inorganic chemists, who analysed reactions in small well-defined complexes in solution. Recently, several groups of biochemists and inorganic chemists (or bio-inorganic chemists) decided therefore that a thorough attack on the problem was necessary to show the constraints on through-protein electron hopping in solution systems. The protein cytochrome *c* is the major experimental object. Different approaches to the problem include protein to protein electron transfer; free small molecule (ferrocyanide) to protein



a, The fold of cytochrome *c*. Black spots, residues known to be affected by redox state changes. The larger effects are towards the bottom and left corner. The major change Met 80 → Ala 85 on substitution at the iron (in *b*) by CN⁻. Other residues known to be affected are shown as black spots. Grosser effects are towards the bottom left-hand corner. These diagrams are of cytochrome *c* from tuna. The site labelled W33 in *a* is equivalent to His 33 in horse cytochrome *c*, the species studied by Isied and co-workers. His 33 is the bound site of electron-transfer ruthenium reagents and is in a conformationally stable region. Lys 13 and 72 are on the front protruding edge of the protein by the haem where reagents such as [Fe(CN)₆]⁴⁻ bind.

transfer; and bound (crosslinked) small molecules (usually ruthenium complexes) to protein electron transfer. The last method, in avoiding the need to determine independently the binding constants and sites of the reagents, is particularly attractive, and it is this method that Isied and collaborators have used to produce their disturbing results.

Before discussing the results of Isied's group in detail, we will briefly describe the ribbon structure of cytochrome *c* (see figure). X-ray diffraction studies of crystals of the protein and nuclear magnetic resonance studies of the protein in solution leave no doubt that the ribbon structure undergoes only a small change from the Fe(II) to the Fe(III) state. Nuclear magnetic resonance additionally provides a dynamic map of fluctuations and their energetics in the protein and has also been used to analyse the structural change and energetics associated with substitution at the iron (see figure). This is a first-order groove-opening reaction dependent on the breaking of the Fe(III)–methionine 80 bond and some small protein rearrangement involving a concerted movement of the section of the protein mainly under the haem. As a result of these studies, the following position was firmly established before the report of Isied's group.

(1) Electron transfer is a reversible phenomenon, which is axiomatic for self-exchange. Provided differences in redox potential are taken into account for forward and back reactions, there is no doubt about reversibility of electron transfer to or from free anions (ferrocyanide), bound cations (ruthenium derivatives) or other proteins (cytochromes *b*). There were no known gated bimolecular reactions involving cytochrome *c*.

(2) A small activation energy is associated with electron transfer for all three types of reaction with some small protein rearrangement (see figure), but electron transfer could not involve a step as energetic as the Fe–Met bond break or ring flips — that is, electron transfer has an activation energy of 10–20 kJ and not of the order of 50–100 kJ.

(3) There is an understandable distance dependence for electron-hop transfer with a rather higher transmission coefficient than expected from theory. This helps to allow long-range (10–15 Å) electron transfer at rates as fast as is commonly found in protein-to-protein electron hopping. Protein–small molecule (free or bound) electron transfer is usually somewhat slower and complicated by solvent relaxation energies, that is, the same type of relaxation, but larger, as that of the small conformation change in the metal protein on redox change, where the protein is the effective solvent.

(4) Electron transfer at these low redox potentials has little dependence on the chemical (amino acid) between the

Dame Honor Fell FRS (1900–1986)

DAME Honor Fell FRS, who died on 22 April, was the pioneer of the technique of organ culture by which small organs or organ rudiments can be successfully grown *in vitro*. In such cultures the various tissue components and their spatial relationship and function are preserved so that the explanted organ closely resembles the parent tissue *in vivo*. In her hands the method proved particularly suitable for the study of developmental processes in fetal organ rudiments but has since been adopted and is still used for the cultivation of adult tissues.

Using this method she made a major contribution in the elucidation of the processes involved in skeletal development. She showed that both avian and murine limb bone rudiments have a remarkable capacity for self-differentiation and will grow and differentiate in the same manner *in vitro* as the organs *in situ*. She further showed that both growth and differentiation could be modified by extrinsic factors, such as vitamins and hormones.

An important example of this modification is the influence of vitamin A on the differentiation of cartilage and ectoderm. She showed that vitamin A added to the culture medium of avian ectoderm modified the direction of epithelial differentiation. The vitamin suppresses normal keratinization and induces mucus-secreting ciliated epithelium instead associated with an increase of sulphur uptake by the metaplastic cells. In embryonic cartilage the vitamin causes a severe breakdown of the matrix with a loss of metachromasia.

This finding formed an important link with her more recent investigations into

the mechanisms involved in joint damage in arthritis. To simulate the rheumatoid pannus she grew pig articular cartilage in association with synovium and showed that the synovium, either in direct contact with the cartilage or in medium conditioned by synovial cells, has, like vitamin A, a deleterious effect on the cartilage ground substance, leading to a loss of proteoglycans and collagen, and of metachromasia.

Together with her colleagues at the Strangeways Research Laboratory, Cambridge, United Kingdom, she traced this effect to a factor secreted by the synovial cells, now identified as the polypeptide catabolin/IL1. Her finding constituted a major advance in the understanding of joint destruction in arthritis and forms the basis of much vigorously pursued international research in this field.

Honor Fell obtained her PhD in Zoology at Edinburgh University in 1923 and the same year joined Dr T.S.P. Strangeways at what was then the Cambridge Research Hospital. After his death in 1926 she became Director of the Laboratory now renamed Strangeways Research Laboratory in 1928, a post she held until 1970.

Under her directorship the small laboratory grew to a major institute of world-wide reputation. In the post-war years the laboratory became a focal point for visiting biologists from abroad whom she trained in *in vitro* methods and who came to rely on her advice, guidance and encouragement, which were unstintingly given. She was one of the most outstanding biologists of this century and her death means the end of an era in biomedical science.

Ilse Lasnitzki

centres, whereas at high potential aromatic groups can act as hopping stations.

The experiments of Isied and collaborators throw all of these conclusions into question. Their results seem to show that the electron-transfer rate, properly corrected for redox potential, can be dependent on the *direction* of electron transfer, that is, it is irreversible. This means that different micro-states are involved for the reaction Ru(II) to Fe(III) and Fe(II) to Ru(III) in cytochrome *c*. Unless there is a gross experimental error in the characterization of species (for example, in structure) this means that this electron transfer can be gated.

The authors offer explanations as to how this could come about and they refer to the conformational states of cytochrome *c* we have described above. It has to be admitted that some states, such as those involved in the flipping of certain aromatic rings and groove openings, do involve activation energies of up to 100 kJ per mole, but it is hard to see how these motions could have anything to do with the electron-transfer steps in only one

direction, especially as they are not involved in the other reactions described above. This is not to say that in some proteins, such as cytochrome P-450, electron-transfer reactions are not gated — in the case of P-450 they are gated by substrate binding. But in the very simple case under discussion we have a puzzle on our hands, in that one set of observations is completely out of line with all the others. Clearly the puzzle can be resolved only with a large series of measurements.

Given the importance of the observations of Isied and colleagues it is essential that at least one other group of inorganic chemists makes the same study independently. Until that is done we face the data with bewildered amazement, frankly hoping that they will go away. If they do not, we must re-examine a considerable part of our thinking about protein energy states and not just about electron transfer. □

R.J.P. Williams and D. Concar are in the Department of Inorganic Chemistry at the University of Oxford, South Parks Road, Oxford OX1 3QR, UK.

Visual observation of lightning propagation

SIR—Reliable scientific observers¹⁻³ have reported discerning direction of movement and propagation in long-duration horizontal lightning flashes. With measured velocities between 5.6×10^3 and 1.1×10^4 m s⁻¹ for horizontal flashes⁴, lightning approaches the upper threshold for detection of movement by the human eye.

During a nocturnal thunderstorm over Ann Arbor on 9 June 1985, we were able to observe and photograph cloud-to-air lightning that appeared to propagate so slowly that we were able visually to observe apparent movement. The photograph of one of these lightning flashes reproduced here neither supports nor contradicts the visual observation of propagation, since it represents a time exposure. Because this photograph shows only a large, widely branched lightning flash, we were prompted to review both the physics of the flash and the limitations of the human visual system to better understand what we perceived.

Even a long-duration lightning flash (≥ 0.4 s) is too brief to allow interpretation of the observation, since the duration of a perceptual experience is around 0.4 s for the dark-adapted visual system⁵, even following the briefest stimuli. The time frame is also too fast to allow the eye to track the flash; so what we must have observed was the temporal ordering of the flash.

A lightning flash usually consists of successive discharges (strokes) along the same channel, and channel sections have been observed to propagate with a mean

pause time of 0.06 s, with about four pause times per horizontal channel⁴. The fovea of the eye is able to determine temporal order perfectly with a pause of 0.3 s between the onset of two targets⁶. Thus the time frame of propagation of long-duration horizontal lightning flashes is more than adequate for the visual processing system to resolve temporal ordering. If the mean duration of such a flash is 0.43 s (ref. 4), the flicker rate of the flash would be around 17 Hz, well below the critical fusion frequency, but too fast for the individual strokes to be counted accurately⁷.

The initial visual stimulus of the lightning flash shown below was probably a single channel extending downwards from the base of the clouds. The image of the initial and subsequent channels would be renewed in the observer's visual system with each stroke of the flash, enhancing the perception of movement and propagation for the observer.

DENNIS G. BAKER

Department of Atmospheric and

Oceanic Science,

University of Michigan,

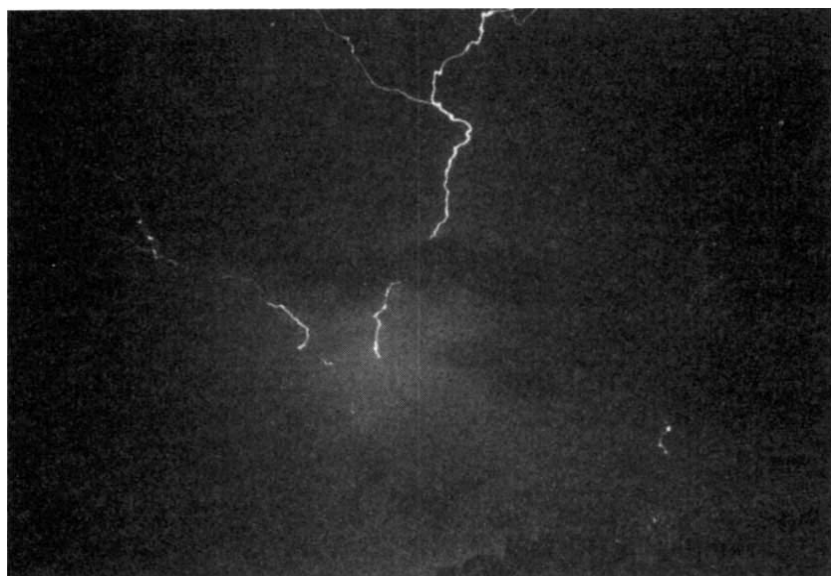
Ann Arbor, Michigan 48109, USA

ANITA BAKER-BLOCKER

Advance Science Consultants,

Ann Arbor, Michigan 48107, USA

1. D'Abadie, A. *Nature* **35**, 342 (1987).
2. Krider, E.P. *Weather* **29**, 24-27 (1984).
3. Brook, M. & Vonnegut, B. *J. geophys. Res.* **65**, 1302-1303 (1965).
4. Brantly, R.D., Tiller, J.A. & Uman, M.A. *J. geophys. Res.* **80**, 3402-6 (1975).
5. Haber, R.N. & Standing, L.G. *Can. J. Psychol.* **24**, 216-229 (1970).
6. McKee, S.P. & Taylor, D.G. *J. opt. Soc. Am.* **A(1)** 620-627 (1984).
7. Ganz, L. *Handbook of Sensory Perception* Vol. 5 (Academic, New York, 1975).



Cloud-to-air discharge emanating from cloud base of about 2 km at 4:30 a.m. EDT on 9 June 1985 at Ann Arbor, Michigan, USA, photographed with a Chinon camera with a 58-mm lens at $f/1.8$, with the shutter held open until after the flash ended, using Kodachrome ASA 25 35-mm slide film.

Controversial glycosamino-glycan conformations

SIR—The issue of the ring conformation of the α -L-idopyranuronate residue in dermatan sulphate, heparan sulphate and heparin — three important biological classes of glycosaminoglycans — was recently debated by Rees *et al.*¹. At least for dermatan sulphate, the authors explained how a postulated conformational equilibrium of the iduronate residue between a prevalent ¹C₄ and a minor ¹C₁ chair form was compatible with experimental observations, including extreme susceptibility of dermatan sulphate to periodate oxidat-

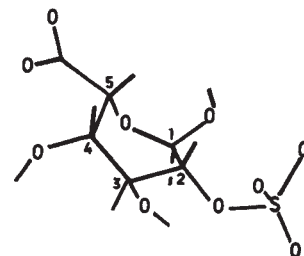


Fig. 1. The ³S₀ skew-boat conformation for the 2-O-sulpho α -L-iduronate residue.

ion. It was stated that "also possible are various distorted chairs and skew boats but these are not discussed here since their consideration would constitute a second order of analysis". This last statement prompts us to air our own views on this controversy.

In a recent work², a detailed force-field study of the conformational characteristics of methyl 4-O-methyl-2-O-sulpho- α -L-idopyranosiduronate demonstrated that the ring may indeed adopt three nearly isoenergetic conformations ¹C₄, ¹C₁ and the skew boat ³S₀. The vicinal coupling constants ³J_{HH} observed for heparin³ and various heparin oligosaccharides containing (I_{2S}), the 2-sulphate iduronate residue⁴, were subsequently interpreted⁵ in terms of conformeric coupling constants computed by using the Karplus-like equation proposed by Altona and colleagues⁶ and the molecular geometry obtained from force-field calculations². For heparin and synthetic oligosaccharide sequences contained in heparin, the I_{2S} residue shows considerable populations of only two conformations, ¹C₄ and ³S₀, the percentage of the skew boat ³S₀ found (40 to 64%) being a function of the sequence. A similar analysis has been achieved for the non-sulphated α -L-iduronate residue in dermatan sulphate (D.R.F., A.P. and M.R., unpublished data). The vicinal coupling constants of the iduronate ring observed in dermatan sulphate⁷ were computed by means of Altona's relationship⁶ relating the ³J_{HH} to the dihedral angles H-C-C-H, which were obtained from force-field calculations on methyl α -L-idopyranosiduronate. The iduronate residue in dermatan sulphate

was shown to be in equilibrium between 1C_4 (58%) and 2S_0 (42%).

We have also analysed data for various natural⁸ or synthetic (J.-C.J., T. Chiba and P.S., unpublished data) glycosides of non-sulphated L-iduronate. A considerable population of the two chair conformers 1C_4 and 2S_0 has been found, in this particular case where the α -L-iduronate residue is unsubstituted at position 4. The main difference between the two sets of coupling constants corresponding to the presence of either 2S_0 or 1C_4 lies in the value of $J_{1,3}$ which in the former case remains small, even when $J_{2,3}$ is quite large.

Finally, 4-unsubstituted glycosides of 2-O-sulpho- α -L-idopyranuronate exist mainly (> 90%) in the 1C_4 conformation as also evidenced by the observation of long-range coupling constants ($J_{1,3}$ and $J_{2,3}$ from 0.5 to 1 Hz) in synthetic⁹ as well as disaccharides¹⁰ and oligosaccharides (B. Perly and M.P., unpublished data) derived from heparin.

The participation in the conformer equilibrium of α -L-iduronate residue located inside a glycosaminoglycan chain of the well-defined² 2S_0 skew boat, which results in a quasi-equatorial orientation of its hydroxyl groups, is an equally plausible explanation for the observed susceptibility to periodate oxidation of unsulphated L-iduronate residue in glycosaminoglycans.

Moreover, the unique conformational flexibility of α -L-iduronate residues may explain some important binding and related biological properties of dermatan sulphate, heparan sulphate and heparin.

B. CASU*, J. CHOAY†, D.R. FERRO‡,
G. GATTI§, J.-C. JACQUINET||,
M. PETITOU†, A. PROVASOLI‡,
M. RAGAZZI‡, P. SINAY|| & G. TORRI*

* Istituto di Chimica e Biochimica
"G. Ronzoni",

via G. Colombo 81,
20133 Milano, Italy

† Institut Choay,
46 Avenue Théophile Gautier,
75782 Paris Cedex 16, France

‡ Istituto di Chimica delle
Macromolecole del CNR,

via E. Bassini 15/A,
20133 Milano, Italy
§ Bruker Spectrospin, Via Pascoli 70/3,
20133 Milano, Italy

|| Laboratoire de Biochimie Structurale,
U.A. 499, U.F.R. de Sciences
Fondamentales et Appliquées,
rue de Chartres, B.P. 6759,
45067 Orléans Cedex 2, France

1. Rees, D.A., Morris, E.R., Stoddart, J.F. & Stevens, E.S. *Nature* **317**, 480 (1985).
2. Ragazzi, M., Ferro, D.R. & Provasoli, A. *J. Comput. Chem.* **7**, 105–112 (1986).
3. Gatti, G., Casu, B., Homer, G.K. & Perlin, A.S. *Macromolecules* **12**, 1001–1007 (1979).
4. Torri, G. *et al. Biochem. biophys. Res. Commun.* **128**, 134–140 (1985).
5. Ferro, D.R. *et al. J. Am. Chem. Soc.* (in the press).
6. Haasnoot, C.A.G., De Leeuw, F.A.A.M. & Altona, C. *Tetrahedron* **36**, 2783–2792 (1980).

7. Gatti, G., Casu, B., Torri, G. & Vercellotti, J.R. *Carbohydr. Res.* **68**, C3–C7 (1979).
8. Sanderson, P.N., Huckerby, T.N. & Nieduszynski, I.A. *Glycoconjugate J.* **2**, 109–120 (1985).
9. Jacquinet, J.-C., Petitou, M., Choay, J. & Sinay, P. *XIIIth Intern. Carbohydr. Symp.*, Utrecht, 68 (1984).
10. Huckerby, T.N., Sanderson, P.N. & Nieduszynski, I.A. *Carbohydr. Res.* **138**, 199–206 (1985).

Ulcerative rhabdovirus in fish in South-East Asia

SIR—Annual outbreaks of a severe ulcerative disease with high mortalities of wild and pond cultured freshwater fish have been reported throughout South-East Asia since 1980¹. The condition is characterized by the appearance of large, deep ulcers on the body and/or head with varying degrees of destruction of the underlying tissues. Many species are considered to be susceptible but the striped snakehead (*Ophicephalus striatus*), one of the economically most important species, has perhaps suffered the most severe losses. Pollution of the natural waterways and fish ponds with insecticides and herbicides, particularly paraquat, is believed by some workers to be the major cause of disease whereas others consider the nature, distribution and pattern of spread of the outbreaks to be more consistent with an infective condition.

Between October 1985 and February 1986, wild and pond cultured fish showing varying degrees of skin ulceration were obtained from widely dispersed locations in Thailand and Burma and subjected to virological examination. Portions of liver, kidney and spleen were removed from affected fish, homogenized in Hanks' balanced salt solution, clarified by low-speed centrifugation and the supernatants

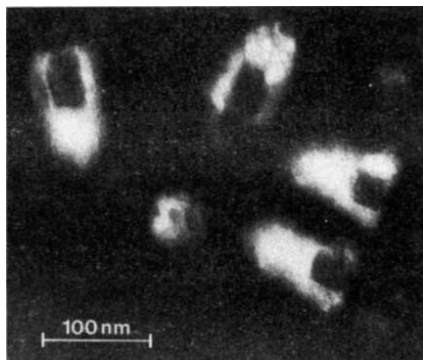


Fig. 1 Rhabdovirus particles from second passage on snakehead cells.

decontaminated by filtration through 450-nm membrane filters. Monolayer cultures of a cell line derived from snakehead fry were prepared using Leibovitz L-15 medium supplemented with 10% fetal calf serum as growth medium and 1 ml aliquots of filtered supernatant added to 25cm² cultures. After allowing absorption for 1 h at room temperature the cultures were overlaid with 7 ml maintenance medium (L-15 + 2% serum) and incubated at 28°C. Cytopathic effects developed between 3 and 14 days post-inoculation. Cells

became granular, rounded up, detached from the monolayer and lysed. This progressed at varying rates over some days until virtual total destruction of the monolayer had occurred leaving only fine, granular debris in the culture medium. Aliquots of media were inoculated onto fresh snakehead cell cultures to confirm transmissibility of the effect.

Electron microscope examination of phosphotungstic acid stained pellets obtained following ultracentrifugation of cell culture fluids from passaged samples revealed bullet-shaped virus particles (Fig. 1). The overall dimensions of these typical rhabdoviruses were 120 ± 10 nm $\times$ 80 ± 5 nm. Morphologically indistinguishable particles were recovered in cell culture from diseased snakeheads from the Bangkok region of central Thailand, Chiang Mai in northern Thailand, Udornthani in northern Thailand and snakehead and freshwater eel (*Fluta alba*) from the Rangoon area of Burma. The size of the particles corresponds closely with that of other recognized fish-pathogenic rhabdoviruses^{2–5}, but further studies are required to determine the relationship of the ulcerative disease isolates to these pathogens.

Although a virus infection has always been recognized as the possible primary cause of disease and virus-like particles of varying morphology have been observed in the tissues of diseased fish⁶, this is the first report of the isolation of a single virus type from more than one species of diseased fish in widely separated geographical areas. However, many diseases of fish cause clinical signs only in adverse environmental conditions and this ulcerative condition seems such a disease. The fact that outbreaks occur in Thailand only during the cooler months of the year suggests that a falling water temperature may be a prime stress factor in this region but other environmental conditions may be precipitating factors elsewhere. Overall we consider the balance of evidence to implicate the rhabdovirus as the likely primary causal agent of the disease.

This work was supported by the UK Overseas Development Administration and FAO. We thank Dr Wattana Wattanavijarn, Chulalongkorn University, Bangkok for facilities and assistance.

G.N. FRERICHS
S.D. MILLAR
R.J. ROBERTS

Institute of Aquaculture,
University of Stirling,
Stirling FK9 4LA, UK

1. FAO Technical Co-operation Project Report 4508 (Department of Fisheries, Ministry of Agriculture and Cooperatives, Bangkok, 1985).
2. De Kinkelin, P., Galinard, B. & Bootsma, R. *Nature* **241**, 465–467 (1973).
3. Cohen, J. and Lenoir, G. *Ann. Rech. vet.* **5**, 443–450 (1974).
4. Hill, B.J., Underwood, B.O., Smale, C.J. & Brown, F. *J. gen. Virol.* **27**, 369–378 (1975).
5. Ahne, W. *Archs Virol.* **48**, 181–185 (1975).
6. Wattanavijarn, W. *et al. Thai J. vet. Med.* **13**, 51–58 (1983); **14**, 31–39 (1984). (In Thai).

Two Cambridge cultures

William Cooper

The Red and the Blue: Intelligence, Treason and the Universities.

By Andrew Sinclair. Weidenfeld & Nicolson: 1986. Pp.179. £12.95.

To be published in the United States by Little, Brown.

I THOUGHT I couldn't bear to read another book about Burgess, Maclean *et al.* — or even a book in which they appeared as characters. I'm allergic to spies. SNEAKS ALL, I call them; regarding the CIA, KGB and MI5 as having more in common with each other than they have with the rest of us. As for the sneaks themselves, some of them are so eaten up with the vanity of sneakery that they can't resist sneaking for both sides at once! Vain and trivial people. Were the CIA, KGB and MI5 disbanded tomorrow, the difference in our world ten years hence would be insignificant.

Burgess, Maclean, Philby and Blunt appear throughout as characters in Andrew Sinclair's *The Red and the Blue*, to which he gives the more explanatory subtitle, "Intelligence, Treason and the Universities". I found I could read the book with admiration and approval. The universities in question are chiefly Cambridge and the period covered begins with the First World War and ends in the present day; the central feature of that period in Cambridge being the glorious era of physical discovery in the Cavendish Laboratory, discovery in atomic and nuclear physics which had a revolutionary impact on world history. The glory was at its peak in the years around 1932 — just when Burgess, Maclean, Philby and Blunt embarked on their private pettifogging tricks that have so hypnotized the British popular press. Mr Sinclair, however, judges his perspective on the whole Cambridge scene with a combination of independence, accuracy and understanding. (He took a double first in History at Trinity College, Cambridge.) His interpretation owes much to the concept of "The Two Cultures", first publicized by C.P. Snow in 1959 — which is when Mr Sinclair graduated, so he only arrived on the scene in person ten years after the half-way mark in his story.

From the end of the First World War to the beginning of the second, the scientific culture is epitomized, obviously, by the Cavendish Laboratory — Rutherford, Chadwick, Blackett, Cockcroft, Rutherford's remarkable Russian protégé, Kapitsa, and others. Kapitsa founded a very influential *open* club, where scientists could publicly exchange information about their discoveries. The literary culture is epitomized by a *secret* society called The Apostles, "a self-electing élite of

those who considered themselves the brightest and best in contemporary literature and philosophy", its members drawn from the privileged classes, male and mostly from Eton, no scientists — Keynes, G.E. Moore, Wittgenstein, E.M. Forster, Bertrand Russell and others. They were given to passionate mutual admiration; devotion to the growth of private virtues (in natural reaction against the collective madness of the earlier war); personal liberation in pursuit of beauty, love and truth — and furthermore, pursuit of beautiful young men, known to them in its platonic form as The Higher Sodomy... oh dear! It is a truth universally unacknowledged that extremely clever people can be ineffably silly: high intelligence is not automatically bonded with commonsense. The Apostles had their Bloomsbury offshoot, known as the Bloomsberries: Lytton Strachey, Leonard and Virginia Woolf, Clive Bell and Vanessa Bell and Duncan Grant, and others. Mr Sinclair takes his stand thus:

In terms of fundamental discoveries that would change our understanding of the structure of the world, the interchanges at the Kapitsa Club made the *conversazione* of the Society of the Apostles seem small beer. Science and technology derived from it would transform human societies, while the military application of atomic research would disturb international relations. Any political or social changes brought about by members of the Apostles could only be trivial. Yet the Apostles had the gift of the pen and the power of the word. They backed into publicity. Their clandestine doings became famous. Along with associates in Bloomsbury, rarely in the field of human endeavour has so much been written about so few who achieved so little. As for the Kapitsa Club, rarely in the field of human discovery has so little been written about so few who achieved so much.

In rebelling against the conventions of the past generation and looking inward with admiration upon themselves, the Society of the Apostles became vulnerable to invasion by Communism during the thirties, when the liberal-minded classes, if not some of the privileged classes, were very properly reacting against Nazism and Fascism. Burgess, Maclean, Philby and Blunt were Apostles converted to active Communism — a secret élite within a secret élite... Mr Sinclair traces their careers, through rightly being overt Soviet supporters when Russia was a major ally of the British during the war, to "moles" working their way, unvetted by virtue of

Apostolic cachet, incredibly into British Intelligence when the West reverted to anti-Communism and the Cold War. Mr Sinclair estimates the military damage they did as trivial, exceeded by the damage done by Fuchs and Pontecorvo, "whose actions only accelerated inevitable discoveries by months or the odd year". More serious was the damage they did to the course of research in those sciences that were of interest to the politicians and the military in prosecuting the Cold War and the armaments race.

Mr Sinclair takes the point that until the tide turned with the advent of war, research was "open", its findings publicly available without restriction: the absence of restriction was, and still is, regarded by scientists as essential to the speediest and most effective advance of their research. There was nothing secret about the discussions of the Kapitsa Club, and when Kapitsa was forcibly recalled to Russia in 1934 he took with him the common assumption that co-operation between his institute and the Cavendish would continue. That assumption was steadily eroded as the politicians and the military on both sides laid hands on nuclear physics; and it was in effect destroyed altogether in 1945 at the bombing of Hiroshima and Nagasaki: as Oliphant put it, "This has killed a beautiful subject". Mr Sinclair adds, "Fuchs's conviction on charges of espionage, followed by the defections of Burgess and Maclean, further soured the possibility of scientific collaboration. Espionage as usual had poisoned reciprocity. Treason was the enemy of reason." Well said! The story is a tragic one; the nature of science permanently marred by its use for unthinkable slaughter; human endeavour brought low by human weakness.

To end on a lighter note, let me earn my fee as a critic by finding fault. Occasional looseness in writing, for example "by months or the odd year". Occasional lapses of judgement, for example equating the damage of Fuchs and Pontecorvo — Fuchs actually was a nuclear physicist, whereas Pontecorvo's speciality was cosmic rays. Occasional failures in proof-reading, for example "Sir Wallace Aken" was Sir Wallace Akers. For my taste too many names per inch of text — the index is a veritable list of All Those People You Would Like To Have Known But Never Had The Chance To Meet. And I began to feel that Mr Sinclair was tiring, his thinking becoming more dispersed, towards the end — when both cultures were themselves becoming more fragmented, more dispersed. Such fault-finding I simply dismiss in the face of the book's major achievement. □

William Cooper, Savile Club, 69 Brook St, London W1, is the pseudonym of the novelist Harry Hoff. Methuen recently republished his book *The Struggles of Albert Woods*.

Observing arms control

Dennis Fakley

Verification: How Much is Enough? By Allan S. Krass. Taylor & Francis: 1986. Pp.271. £25, \$45.

CURRENT arms control treaties, such as the 1963 Partial Test Ban, 1967 Outer Space, 1972 Sea-Bed, 1972 Anti-Ballistic Missile, 1975 Biological Weapons and 1978 Environmental Modification agreements, contain little or no provision for cooperation between the signatories on measures to verify that the obligations in the treaties are being observed. This lack of verification provisions can be attributed either to the effectiveness of so-called national technical means of verification or to the relative unimportance of the treaties from a security viewpoint. Other treaties, which have been negotiated (but not ratified) or have been discussed — Threshold Test Ban, Peaceful Nuclear Explosions, Strategic Arms Limitations II, Comprehensive Test Ban, Chemical Weapons — are considered to be much more security-sensitive, and thus to require special measures to verify compliance. Western dissatisfaction with negotiated or proposed verification arrangements for these treaties accounts, in part, for the failure to complete them.

Any analyst of arms control verification is faced with a surfeit of American sources and a dearth of Soviet ones. In consequence, it is difficult to maintain a balance between American and Soviet views, upon which a verification regime has to depend. Professor Krass quotes extensively from American sources on both American and Soviet positions — without recognizing that the latter may be inaccurately reflected — and finds ample support for his conclusions.

In the Introduction, verification is taken to be the action of demonstrating compliance with treaty obligations. However, an arms control treaty usually requires the parties concerned to refrain from some specified activities rather than to undertake them; hence, verification has to be concerned with demonstrating an absence of non-compliance. As Professor Krass observes, a negative cannot be proved, but he could have pointed to the possibility of developing measures to provide sufficient assurance that treaty obligations were not being flouted.

The first main chapter, "Technology", is an account of the monitoring capabilities of satellite photography, ground- and satellite-based radars, seismology, nuclear explosion detectors, electronic signals interceptors and safeguards systems of the type operated by the Inter-

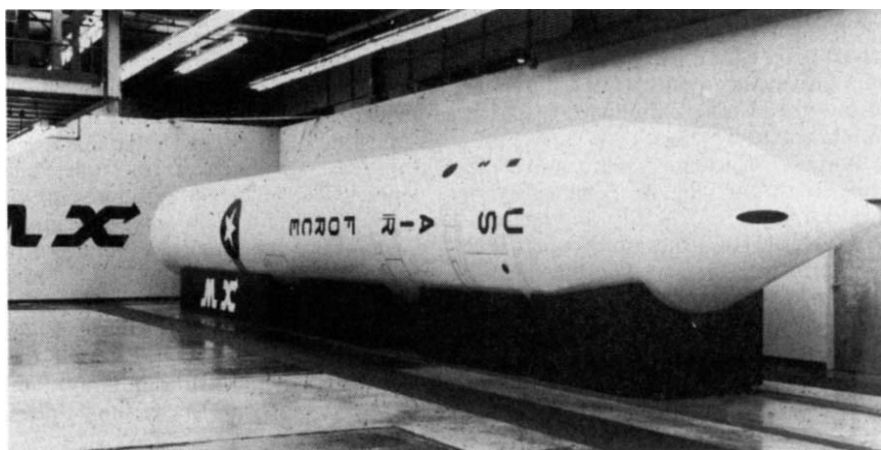
national Atomic Energy Agency (IAEA). While reference is made to some of the problems of achieving the theoretically available performances from these systems, such difficulties are largely discounted in the conclusion that a substantial number of arms control bans or limits could be verified with a high degree of reliability. This would not be universally endorsed.

In the next chapter, "Politics", Professor Krass outlines two major differences between the American and Soviet political approaches to arms control verification. He contrasts the American requirement, as a negotiating precondition, for assurance on the verifiability of an arms control measure, with the Soviet requirement for a definition of the measure to be negotiated before the question of

confidence that any American attempt to circumvent a treaty would inevitably become public knowledge — and sooner rather than later.

American bureaucracy certainly complicates the process of arms control negotiation, and the separation of executive and legislative functions means that Presidential proposals are not necessarily enacted. This would not, however, appear to warrant Professor Krass's suggestion that verification provisions should be made less technically precise to avoid the risk of misunderstanding in Congress. On the contrary, precision is essential if the scope for genuine differences in interpretation is to be minimized.

It is undoubtedly true that the assessment of whether a verification regime is adequate depends on the prevailing politi-



Full-scale mockup of the MX intercontinental ballistic missile developed by the US Air Force

verification is discussed. He also compares their respective positions on on-site inspection, suggesting that they may be held for not entirely honourable reasons. He is, perhaps, over-optimistic about these differences becoming smaller.

Professor Krass appears to underestimate the importance of the fact that the closed nature of Soviet society militates against American confidence in treaty compliance. The American authorities recognize that, if the Soviet Union decided that it would be worthwhile violating an arms control obligation, it would not be constrained by public opinion. While not necessarily believing that the Soviet Union would deliberately enter into an agreement with the intention of disregarding its obligations, the possibility that it might do so leads the Americans to presume that treaty violations would occur unless they were made too unattractive by the verification provisions of the treaty. It is no argument to claim that the American intelligence services have been successful, in the post-war era, in penetrating the screen of Soviet military security. The successes are known but the failures may not be apparent until it is too late. On the other hand, the Soviet Union, no matter what it says, can be reasonably

cal climate. This leads Professor Krass to conclude that unreasonable demands should not be made of the verification arrangements lest they become unacceptable with a change of political opinion. His conclusion is consistent with the underlying theme of the book, i.e. that arms control measures are needed so urgently that even risky compromises are justified if they lead to a negotiated agreement. He does not offer the option of postponing arms control agreements until robust verification measures become available.

It is highly questionable whether any real and stable arms control agreement can be reached as long as there is the present lack of trust between East and West, and Professor Krass rightly questions whether verification arrangements would build or undermine international confidence. He appears to believe that it needs a certain leap of faith to generate a modicum of trust between the superpowers to serve as a nucleus on which confidence could grow. Otherwise, verification arrangements are likely to reinforce existing prejudices. This is consistent with the view that it is futile to seek to resolve political differences between East and West by purely technical means; at best, they can only contribute to a pro-

cess which has been politically initiated.

In the third and final main chapter, "Technology and Politics", Professor Krass attempts to correlate the technical and political issues he has raised. He examines the legitimacy of various verification technologies and sees the possibility of friction arising from the lack of an adequate international legal and institutional framework to regulate them. He also discusses how to deal with real and mistaken assessments of treaty non-compliance, and recommends that charges should be pursued vigorously only in cases where the suspected violation could have military significance; the political implications of breaches should largely be disregarded to avoid a threat to the continuation of the treaty. This would demand a tolerance which many might consider unrealistic. It would have been interesting to have had a deeper analysis of the different attitudes of the United States and the Soviet Union towards treaty compliance, and a recognition that "one man's compliance could be another's violation".

On-site inspection has for long been a bone of contention in arms control negotiation. Professor Krass argues that inspection rights have in some cases been sought for other than technical reasons, and that the effectiveness of on-site inspections in many arms control contexts has not been established. He predicts that there are insurmountable obstacles to any agreement on mandatory rights of inspection at other than pre-determined locations and that any inspection rights will be granted only when there is a reasonable degree of trust between parties to the treaty. Irrespective of the technical case for on-site inspection, politically it has become one of the measures by which the intentions of the Soviet Union will be judged.

Professor Krass deals only briefly with multilateral arms control but he would welcome an internationalization of verification. Here he fails to recognize that a sovereign state cannot possibly risk delegating to an international body responsibilities for activities that might impinge on its fundamental security interests. The safeguards regime operated by the IAEA does not provide a precedent.

If readers expect to find quantitative answers to the question posed in the title of this book, they will be largely disappointed. Professor Krass does, however, provide much relevant information in an easily digestible form and, so long as the reader recognizes the viewpoint from which he writes, the book will serve as a useful introduction to some of the issues of arms control verification. □

Dennis Fakley, 14 Coval Gardens, London SW14 7DG, was formerly Assistant Chief Scientific Adviser (Nuclear) at the Ministry of Defence, and was Deputy Leader of the British Delegation during the 1977-1980 Comprehensive Test Ban negotiations. The opinions expressed in this review are his own.

Committee ecology

Michael J. Crawley

Ecological Knowledge and Environmental Problem-Solving: Concepts and Case Studies. Committee on the Applications of Ecological Theory to Environmental Problems, chairman Gordon H. Orians. *National Academy Press, Washington DC: 1986. Pp. 388. Pbk \$24.50. Distributed in Britain by Wiley £28.60.*

This is a curious book. Its origins lie in the recognition, by the Board on Basic Biology of the United States National Research Council's Commission on Life Sciences, that it should be concerned with the basic biology of ecosystems. The first aspect of ecology to receive the Board's attention was the current status of ecological concepts, and their applicability to specific environmental problems. Following a workshop in 1979, a committee was established under the chairmanship of Gordon Orians. This book represents the fruits of its labours.

There are two parts to the report. The first attempts to encapsulate "ecological knowledge" as it relates to environmental problem-solving. The second consists of 13 case-studies, selected to illustrate the use of ecological information in dealing with a wide variety of environmental problems. The intended audience is "those who prepare, receive, and use environmental evaluations and management plans; legislators and regulators concerned with environmental issues; and teachers and students in the natural and environmental sciences". Given this audience, the section "Kinds of Ecological Knowledge and Their Application" in the first part of the book has to be judged as unsatisfactory; it varies between being unacceptably shallow and downright misleading. A great many of the statements, though undoubtedly true (and probably both worthy and important), are so brief as to be unintelligible. Take this, for example: "many ecological interactions are characterized by strong non-linear effects brought about by slight changes in key factors". One would need a thorough grounding in the theoretical ecology to get the real message.

A number of specious fallacies also get recycled in this part of the book. For example, on p. 28, we read that "the greatest number of individuals of a species that the environment can support is called the carrying capacity (K)". This is wrong; K is the *equilibrium* population size in models such as the logistic. In discrete-time versions of the logistic, overshoots of K are possible and perfectly reasonable, representing population sizes temporarily in excess of equilibrium. Again, on the same page, we find this gem: "Many pest

species, such as weeds, are poor competitors". Tell that to any gardener! Weeds are *exceptionally* competitive; rather, the point here is that many of them are short-lived, and also are rather unlikely to recruit from seed under shady conditions.

The second half of the book is much better, and the case-histories are carefully chosen and extremely interesting. They include well-known examples such as the biological control of California red scale, the Garki project on malaria control in West Africa, management of the North Pacific halibut fishery and the environmental effects of DDT. More recent studies include research on nuclear radiation, eutrophication, forest felling policy and wildlife management. It would have been good to see more of them, allowing the inclusion of topics such as acid rain, desertification, release of genetically engineered organisms, the regulation of human population growth and so forth.

At the end of each case-study is a kind of headmaster's report by the editors. These "Committee Comments" are one of the highlights of the book, and have a delightfully superior tone about them — "relevant ecological theories, for the most part, proved inspirational or didactic, rather than directly applicable", or "it is interesting that the scoping of this project (and in fact, most of applied ecology) is somewhat myopic"! Another unintended bonus lies in a liberal sprinkling of parenthetical one-liners of the utmost inscrutability. For instance: "(if plants can be regarded as predators on photons)"; "(all predators must eat regularly)"; "the guts of herbivores are (technically outside their bodies, inasmuch as no cell membrane has to be crossed)"; or "(all CO_2 molecules can be treated as identical)"!

The recommendations that emerge, however, are laudable. In carrying out environmental projects, managers are urged to: (1) involve scientists from the beginning; (2) treat projects as experiments; (3) publish information; (4) set proper boundaries on projects; (5) use natural history information; (6) be aware of interactions; (7) be alert for possible cumulative effects; (8) plan for heterogeneity in space and time; and (9) prepare for uncertainty and think probabilistically. This scheme would have made an excellent plan for the first section of the book, where the ecological principles underlying each of the recommendations could have been explained in context. As things are, however, the first half of the report is as weak as the second half is strong, and the book stands as a monument to the pitfalls of committee authorship. □

Michael J. Crawley is a Lecturer in Ecology at Imperial College, Silwood Park, Ascot, Berkshire SL5 7TY, UK.

AIDS without emotion

Peter Newmark

Mobilizing Against AIDS: The Unfinished Story of a Virus. By Eve K. Nichols. *Institute of Medicine/National Academy of Sciences/Harvard University Press: 1986. Pp.198. Hbk \$15; pbk \$7.95. To be published in Britain on 18 August, hbk £12.75, pbk £6.75.*

IT HAS become commonplace to describe as astounding the rate at which research on the AIDS virus has proceeded since the first isolation of the virus at the Pasteur Institute just over three years ago. The pace continues, so that the writer of any book on the subject faces an impossible task in trying to be fully up to date.

Although, in theory, handicapped in this regard by basing her book on a conference held in 1985, Eve Nichols has been able to add new information as recent as this April. Nonetheless, she has missed out on key new discoveries including that of a new gene in the virus and of what seems to be a second AIDS virus in some Africans. This just goes to prove the point about the rate of research, which is matched only by the rate of publication of books on AIDS. But Eve Nichols seems to have performed as good a job as anyone in writing a sensible account for an educated public.

The meeting to which the book owes its origins was the annual conference of the Institute of Medicine in the United States. There were about a dozen speakers, including both Luc Montagnier of the Pasteur — the only non-American — and Robert Gallo of the National Institutes of Health. Each of the eight chapters in the book is based on a few of the talks, and the scientific, medical and public health aspects of AIDS are all dealt with in a straightforward and unemotional way.

Care is taken to tread a middle course between the pessimists and the optimists on issues such as the rate of spread of AIDS in heterosexual populations and the prospects of effective vaccines or therapy. Adequate attention is drawn to such

thorny questions as whether AIDS should be made a notifiable disease, so that contacts could be traced, when guarantees of confidentiality seem to have a way of being less than cast iron and when there is growing discrimination against even healthy carriers of the virus by employers and insurance companies.

The book comes complete with a glossary and a list of organizations (only American) to contact for help and information. Also included, as appendices, are the Public Health Service recommendations on preventing transmission in the workplace, care of infected children and reducing the risk of transmission. There is additional advice on how to prevent transmission, but none on how to put into practice one of the recommendations:

"Everyone should know the infection status of his or her sexual partners; those who are not infected should avoid intercourse with those who are infected".

Were this not a book aimed at the public, and particularly the public of the United States, it would be more fair to criticize the lack of recognition of some of the European contributions to the understanding of AIDS. In the same vein, it is a pity that less than two pages are devoted to the exceedingly serious problems of AIDS in Africa. But then, does the American public at large care about African AIDS? And for that matter, how many African countries would welcome a book such as this which dealt with their own AIDS problem for their own people?

Peter Newmark is Deputy Editor of Nature.

Nimble electrons

David R. Rosseinsky

Hopping Conduction in Solids. By Harald Böttger and Valerij V. Bryksin. *Akademie-Verlag, Berlin/Royal Society of Chemistry: 1986. Pp.398. DM 140, £44.*

"LOVELY 'OPS", murmurs the drill sergeant quaffing a pint of beer in the television advertisement, while watching a platoon of recruits hop past him. So it is with electrons, which in certain solids move, and conduct, by hops from site to site. The other kind of electron motion, within the extended bands allowed by multi-centre wave functions pervading condensed matter, has curiously enough never acquired a complementary term like "swanning" or "floating" or even "shining": it is simply band conduction. Metals have it anyway, but crystalline semiconductors have bands only of high energy separated by a gap from the states populated by the bulk of electrons, which requires an energy input to mobilize the electrons. Most of the vast structure of solid state physics encompasses band systems, and the recent interest in hopping arose largely from the fact that amorphous semiconductors have irregular structures generating sites which are envisaged to lie within the gap. As these are more or less discrete from the bands, and each other, so are the motions of visiting electrons, hence the designation "hops". Nevill Mott's 1977 Nobel prize was in part awarded for his studies of amorphous semiconductors, and those who in the 1970s based their learning on Austin and Mott, and Mott and Davis, will be avid readers of the text at hand.

Professor Böttger in Magdeburg and Dr Bryksin in Leningrad have themselves contributed substantially to the subject and here provide a comprehensive account of the theory of the hopping mech-

anism. After elaboration of Hamiltonians for the hopping system, the small polaron model is described, in which the charge in question, together with the self-entrapping reciprocal effect on its ambience, represent the moving entity. The theoretical framework is then applied to observations on disordered systems. Ionic and atomic hopping are addressed next, and the final chapter is an up-dating of the material to mid-1985, inserted in proof as an addendum to the 1983 manuscript. This is perhaps not an entirely comfortable addition but it represents an excellent way of circumventing publication delays, fulfilling the aim of the series, of which this book is a member, to provide quick dissemination of information. Other authors and publishers could profitably follow suit. Part of the commendable immediacy has apparently been achieved by direct authorship without the intervention of a translator. Nearly a third of the references are those added in proof, which is indicative of the recent growth of interest in the mechanism.

If the book is somewhat austere in aspect it is quite readable withal. Few, however, who are not involved in the minutiae of conduction or electron transfer within solids will approach it lightly, but, besides physicists, solid state chemists will increasingly find the need for recourse to one or other aspect of the exposition. Site-to-site electron transfer occurs in (chemical) mixed-valent solids where the sites are regular and rigorously defined, but such promising, even ideal, routes towards verifications of the theories have not yet entered the province of the monograph. Biological electron transfer between donor and acceptor sites is of the same kind as that considered here, and so the ramifications of this quite complex motion will extend across the sciences — perhaps by discrete hops. □

David R. Rosseinsky is a Reader in the Department of Chemistry, University of Exeter, Exeter EX4 4QD, UK.

• *AIDS: Papers from Science, 1982 – 1985* is a collection of 108 research papers and reports, edited by Ruth Kulstad, that have appeared in *Science*. Arranged chronologically, the papers show the progress in AIDS research and indicate the directions in which it might go. An introduction by Dr Myron Essex, chairman of the Department of Cancer Biology at the Harvard University School of Public Health, provides an overview of AIDS research. Available from the American Association for the Advancement and Science, Marketing Dept. A, 1333 H St NW, Washington DC 20005, priced Hbk \$32.95, pbk \$19.95 plus \$1.50 p+p.

Geochemical constraints on core formation in the Earth

John H. Jones & Michael J. Drake

Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721, USA

New experimental data on the partitioning of siderophile and chalcophile elements among metallic and silicate phases may be used to constrain hypotheses of core formation in the Earth. Three current hypotheses can explain gross features of mantle geochemistry, but none predicts siderophile and chalcophile element abundances to within a factor of two of observed values. Either our understanding of metal-silicate interactions and/or our understanding of the early Earth requires revision.

CORE formation has been described as perhaps the single most important thermal event in the history of the Earth¹. This description is certainly true if the core formed by segregation of metal in an initially homogeneous body the size of the present Earth, as the heat liberated in such an event would raise the temperature of the Earth by ~2,000 °C (ref. 2). However, less extreme models in which core formation is concomitant with accretion appear to be more probable³. Regardless of whether core formation resulted in catastrophic changes in the Earth, better knowledge of the core-forming event would certainly elucidate important aspects of the histories of the Earth and Moon⁴.

Geochemists have traditionally divided the chemical elements into four classes: (1) lithophile—those elements which prefer to be combined with oxygen and which typically reside in rocky material; (2) siderophile—elements which prefer to be in the metallic state; (3) chalcophile—elements preferring to be associated with sulphur; and (4) atmophile—elements so volatile that they are typically concentrated in a planet's atmosphere. Obviously, membership in one of these classes is not invariant, but is modulated by the ambient conditions. For example, as the temperature increases, elements which normally exist as solids or liquids will vaporize and become gases, and, in fact, the classes of lithophile, siderophile and chalcophile are typically subdivided into groups of elements of similar volatility.

In principle, the record of core formation is contained in the siderophile and chalcophile element contents of rocks derived from the Earth's mantle. Certainly these elements are much more depleted in the upper mantle than lithophile elements of similar volatility, implying that they record a depletion event unseen by the lithophile elements—presumably core formation. In practice, these elemental abundances have been attributed to either gross disequilibrium between the core and upper mantle, or mixing of siderophile- and chalcophile-element-rich material into the upper mantle after core formation, or both^{5,6}. In this view, little information about core formation may be obtained from mantle siderophile and chalcophile elements because their abundances were established either after core formation ceased or as the results of processes which never approached equilibrium. The evidence cited for disequilibrium between mantle and core during core formation is simply that the concentrations of siderophile and chalcophile elements in mantle-derived rocks are orders of magnitude higher than would be expected if these rocks had ever equilibrated with metal⁶. Furthermore, many of the siderophile elements in the Earth's mantle exist in chondritic proportions (that is, with elemental abundance ratios characteristic of primitive meteorites (chondrites) which have not undergone chemical processing). Thus, successful models of core formation and for siderophile and chalcophile element abundances in the Earth's upper mantle must be able to explain both high abundances and the apparently unfractionated ratios of some (but not all) element pairs.

Models of core formation are further complicated by our lack of knowledge of the chemical composition of the core. All models of the Earth's core require the presence of a 'light element' to reduce the overall density³. If the light element is sulphur, silicon or carbon, then these elements were probably dissolved into the metal (which eventually formed the core) at low pressure as the Earth grew by accretion, and may be relevant to understanding the core-forming process. If the light element is oxygen or a mixture of oxygen and lithophile elements which have followed oxygen into the core, then the light element must have been incorporated at high pressure⁴ and may not have played a role in establishing the siderophile and chalcophile element abundances of the upper mantle.

The behaviours of siderophile trace elements in complex, natural metal-silicate systems are poorly known. This lack of knowledge has led us to determine partition coefficients (D_s ; for definition, see below) between solid metal, sulphur-bearing metallic liquid and silicate liquid for a suite of siderophile and/or chalcophile elements under controlled laboratory conditions ($T = 1,250\text{--}1,270\text{ °C}$; $P \leq 1\text{ bar}$). These partition coefficients, together with estimates of the mantle abundances of the same elements, may be used to evaluate three physically plausible but very different hypotheses of core formation. These hypotheses are: (1) inefficient core formation, involving equilibrium between solid and liquid metals and silicates, with a small fraction of solid and liquid metal remaining trapped in the mantle, subsequently to be oxidized⁷; (2) equilibrium between an Fe-S-O metallic liquid and the mantle⁸; and (3) heterogeneous accretion involving a late 'veneer' of oxidized chondritic material which did not segregate into the core^{6,9}. In each of these hypotheses, metal is added to the surface of the growing Earth, and initial interaction between metal and silicate will occur at relatively low pressures. Hence, our partition coefficients, which are obtained at low pressure, may be applicable. Later, we shall return to this point and further discuss the importance of pressure on partitioning equilibria.

A fourth hypothesis, that siderophile and chalcophile element abundances in the Earth's upper mantle were established by partitioning between metal and silicate at very high pressures, perhaps at the core-mantle boundary, will not be explored for three reasons. First, metal-silicate partition coefficients have not been determined at megabar pressures, precluding quantitative evaluation. Second, evidence discussed below suggests that siderophile and chalcophile element concentrations in the mantle were established no later than 3.5 Gyr ago. Thus, material must have been efficiently cycled from the core-mantle boundary to the upper mantle and have been well mixed in <1 Gyr. Such effective mixing appears problematical. Third, it seems unlikely that equilibration of siderophile elements between the core and lower mantle would have preserved chondritic ratios of elements such as Ni and Co, which differ greatly in their siderophile tendencies.

Table 1 Experimental conditions and results

Element	T (°C)	$-\log f_{O_2}^*$	$D_{(SM/LM)}$ (range)	$D_{(LS/LM)}$ (range)	% $S_{LM}^\ddagger$	$N^\S$
W	1,250–1,270	12.3–12.5	20 (15–30)	~ 1 (0.12–1.9)	19–24	3
Re	1,250–1,260	13.0–13.5	51 (51–>60)	5×10^{-4} (5×10^{-3} – 5×10^{-5})	23–24	2
Ir	1,270	12.5–13.2	46 (41–52)	5×10^{-5} (5×10^{-4} – 5×10^{-6})	22–23	2
Mo	1,260	—	2.45	—	25	0
Ni	1,270	12.5	1.24	2×10^{-4} (4×10^{-4} – 1×10^{-4})	23	
Co	1,260	12.6–12.7	2.3 (2.2–2.5)	0.0075 (0.006–0.010)	24–25	2
Au	1,270	12.2–13.0	1.0 (0.97–1.1)	10^{-4} (3×10^{-4} – 5×10^{-5})	20–22	2
P	1,250	12.6	0.28	0.24	18	1
Ga	1,270	12.5	3.5 (2.6–4.3)	0.8 (0.6–1)	19–22	2
Ag	1,250	12.7–13.0	~ 0.01	0.01 (0.01–0.014)	23–27	2
Pb	1,270	12.2	BDL¶	0.15	20	1

* Calculated from the Fe content of the solid metal and the FeO content of the silicate glass using the iron–wüstite data of ref. 41.

‡ Weight per cent sulphur in liquid metal phase.

§ Number of individual experiments which contained silicate melt.

|| All experiments contain Ni but only one sample (the most oxidized sample to receive a high-neutron-flux irradiation) contained measurable Ni in the silicate glass. The range for $D_{(LS/LM)}$ is that for two analyses of the same glass.

¶ Below detection limit.

The first three hypotheses will be explored in more detail. First, however, we will discuss constraints obtained from experimental and natural samples which must be satisfied by any successful model of core formation in the Earth.

Constraints

Partition coefficients. Our experimental partitioning techniques have been described elsewhere¹⁰ and will not be repeated in detail here. In general, samples containing Fe–Ni metal, natural pyrite (FeS₂), synthetic basaltic glass and a tracer element are placed in an alumina crucible and sealed in an evacuated silica tube with a separate Fe–metal–silica assemblage designed to maintain the oxygen fugacity of the system (f_{O_2} , equivalent to the partial pressure of oxygen) near the quartz–fayalite–iron (QFI) buffer (thus ensuring that iron metal is a stable phase). Quenched samples of coexisting solid metal (SM), liquid metal (LM) and liquid silicate (LS) are analysed by electron microprobe. If the tracer concentration in the silicate glass is too low for microprobe analysis, aliquots of glass are separated from metal and analysed by instrumental neutron activation. Table 1 gives experimental conditions and results for W, Re, Ir, Mo, Ni, Co, Au, P, Ga, Ag and Pb—a suite of elements which covers a wide range of nebular condensation temperatures (volatilities) and geochemical behaviour. In this suite W, Re and Ir are the most refractory and Ag and Pb are the most volatile.

Siderophile abundances in the bulk Earth. The presumed extraction of most of the siderophile and chalcophile elements to the core makes calculation of the bulk-Earth concentrations of these elements difficult or impossible. However, on the basis of lithophile elements, which do not readily combine with Fe or S, we infer that the bulk Earth is approximately chondritic in composition. We do not mean to imply by this that the Earth is compositionally identical to the group CI chondrites, but that refractory siderophile elements should be present in chondritic relative proportions and that more volatile elements may be depleted to varying degrees. Figure 1 shows the concentrations of siderophile and chalcophile elements in several types of chondritic meteorites. Refractory siderophile elements are usually present in higher concentrations than in group CI, while more volatile elements are typically depleted. Again, although the concentrations of siderophile and chalcophile elements in the bulk Earth are unknown, it is reasonable to require that models of bulk-Earth abundances of these elements be consistent with siderophile and chalcophile element concentration patterns in chondrites.

Siderophile abundances in the upper mantle. During the past decade, a large quantity of high-quality data on mantle samples (such as lherzolites) and basalts has become available. Morgan

*et al.*⁶ and Jagoutz *et al.*¹¹ have presented convincing evidence that their lherzolite samples (or at least their most 'fertile' lherzolite samples) have not undergone large amounts of partial melting, and hence are most likely to record primitive mantle abundances, at least for the compatible elements. Thus, when possible, we have used the mantle abundance estimates of refs 6 and 11. For the incompatible elements P, Mo and W, abundances are inferred from the trace-element systematics of basalts (see below). These estimates are listed in Table 2.

Mineral–melt partitioning in basalts. The partitioning of Ni and Co and between minerals and silicate melts is moderately well known from laboratory experiments, but these elements are the exception rather than the rule. For most of the elements in our suite, solid silicate (SS)/liquid silicate (LS) partition coefficients for element i ($D_{SS/LS}$) must be evaluated from element correlations in natural basalts. For example, a comparison of data for basalts and ultramafic xenoliths on Re versus Ir and Au versus Ir diagrams from ref. 12 implies that $^{Au}D_{SS/LS}$ and $^{Re}D_{SS/LS}$ are of order unity, but that $^{Ir}D_{SS/LS}$ is ~ 50 . (Note in Table 1 that during core formation, trivially small fractions of Au, Re or Ir enter the silicate phase, making the exact value of these particular silicate partition coefficients of little consequence. Also, although the host phase of Ir during silicate partial melting is unknown, the large differences in partitioning behaviour between Au and Ir strongly imply that the host is an oxide or silicate.)

Drake¹³ has used element i versus La diagrams to make geochemical distinctions between terrestrial, lunar and eucritic basalts. Many siderophile and chalcophile elements, when plot-

Table 2 Siderophile and chalcophile element abundances in the upper mantle

Element	Concentration/CI*	Ref.
W	0.11	15
Re	0.0075	6
Ir	0.0075	6
Mo	0.064	15
Ni	0.19	11
Co	0.21	11
Au	0.02	6
P	0.058	14
Ga	0.3	11
Ag	0.087	6
Pb	0.09†	42

* Concentrations normalized to CI chondrite ref. 43.

† Mean of total Pb and unleachable Pb.

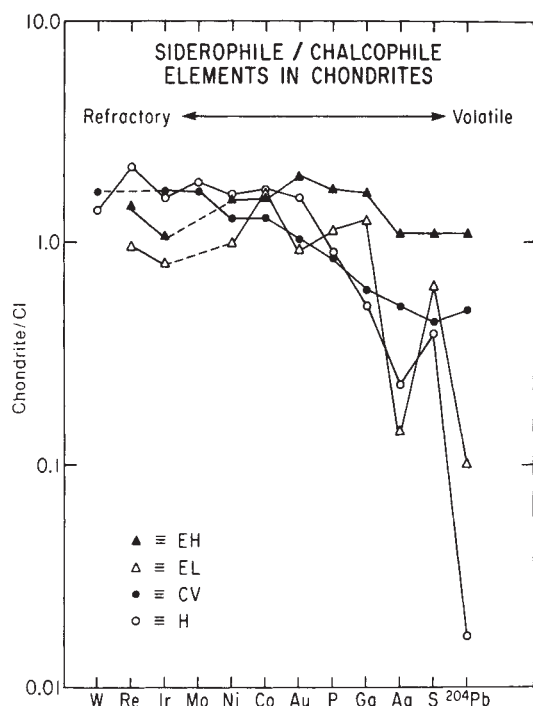


Fig. 1 Siderophile and chalcophile element abundances for four different types of chondritic meteorites ($\blacktriangle$, EH; $\triangle$, EL; $\bullet$, CV; $\circ$, H), normalized to the concentrations of these elements in CI chondrites. Elements on the left are more refractory and elements on the right are more volatile. Relative to CI chondrites, refractory siderophile and chalcophile elements are typically enriched in other chondrite types and volatile siderophile and chalcophile elements are typically depleted. Chondritic materials provide estimates of the siderophile and chalcophile element abundances that may be expected in the bulk Earth (see text). Data are from the following sources: E, chondrites, refs 45 (Re, Ir, Ni, Au, Ag), 43, 46 (Co, Ga), 43, 47 (P), 43, 48 (S), 43, 49 (^{204}Pb); CV, chondrites, refs 50 (Ir, Co, Ni, Au, Ga), 43, 51 (Re, Ag), 43, 52 (W), 53 (Mo), 43, 47 (P), 43, 54 (S), 43, 55 (^{204}Pb); H chondrites, refs 43, 56 (W, Re, Ir, Mo, Ni, Co, Au, P, Ga, S), 43, 57 (Ag), 43, 58 (^{204}Pb). When only one reference is given, both the chondrite group analysis and the CI normalization analysis are reported in that reference. When two references are given, the first is the source of the CI normalization.

ted on such a diagram, form linear trends which yield information on the depletion of element i in the mantle (relative to chondritic abundances) and permit the partition coefficient of i between basalt and mantle residuum to be estimated. For example, Ga concentrations in basalts increase by a factor of two, while La concentrations increase by two orders of magnitude in natural terrestrial samples¹³. If we assume that this basalt trend is formed by differing degrees of equilibrium partial melting of similar mantle sources and that $^{La}D_{\text{SS/LS}} \leq 10^{-2}$, then $^{Ga}D_{\text{SS/LS}} \approx 0.4$. Correlations of Pb with La show almost exactly the same variation as the Ga–La diagram, implying that $^{Pb}D_{\text{SS/LS}}$ is also ~ 0.4 . The good (approximately 1:1) correlation of P, W and Mo with La or another lithophile incompatible element in basalts^{14,15} implies that these elements are all very incompatible in silicate systems. Diagrams of phosphorous, W and Mo versus incompatible elements are also the best means of ascertaining the mantle abundances of these elements^{14,15}. Analyses of Ag in terrestrial basalts are very scarce; however, based on the available data^{16,17}, Ag appears to be weakly incompatible like Ga and Pb. Consequently, we have adopted a value of $^{Ag}D_{\text{SS/LS}} = 0.4$, although it could be larger. For example, the MORB (mid-ocean-ridge basalt) data of ref. 17 are consistent with values of $^{Ag}D_{\text{SS/LS}}$ between 0.1 and 0.5, for 5% equilibrium partial melting.

Summary. Table 3 lists our adopted partition coefficients. As we shall see below, the uncertainty associated with these partition coefficients, while often large, is of the same order as the uncertainties regarding the physical conditions under which core formation occurred. In this sense at least, we feel that zeroth-order modelling, using the data of Tables 2 and 3, is justified. We will now examine hypotheses for core formation in the light of these data.

Hypotheses

Inefficient core formation. The segregation of the core from the mantle is envisaged as an equilibration between four phases, namely, solid silicate, liquid silicate, solid metal and liquid metal, followed by incomplete separation of metal from silicate. Equilibrium between these four phases is described by three sets of partition coefficients (D) for each element at specified temperature, pressure and oxygen fugacity: $D(\text{solid metal/sulphur-bearing metallic liquid})$, $D(\text{liquid silicate/sulphur-bearing}$

Table 3 Adopted partition coefficients

Element	$D_{(\text{LS/LM})}$	$D_{(\text{SM/LM})}$	$D_{(\text{SS/LS})}$
W	1	36	0.01
Re	5×10^{-4}	83	1
Ir	5×10^{-5}	83	50
Mo	$8 \times 10^{-4*}$	2.45	0.01
Ni	2×10^{-4}	1.33	10†
Co	7×10^{-3}	2.3	3†
Au	1×10^{-4}	1.3	1
P	0.24	1.7	0.02
Ga	0.8	6.0	0.4
Ag	0.01	0.01	0.4
Pb	0.15	0	0.4

* Calculated from our data and that of ref. 25.

† Ni and Co $D_{(\text{SS/LS})}$ values estimated from ref. 44; all others estimated from natural basalt systematics (see text).

metallic liquid) and $D(\text{solid silicate/liquid silicate})$ where $^{i}D_{(\alpha/\beta)}$ is defined to be the weight concentration of element i in phase α divided by the weight concentration of element i in phase β .

The partitioning behaviour of the elements in the experimental suite can be generalized in the case of this hypothesis. Mantle concentrations of Au, Ir, Mo, Ni and Re are determined almost solely by the amount of trapped metal phases. Silver is concentrated nearly exclusively in the trapped metallic liquid; P, W, Ga, Co and Pb are intermediate in their siderophilic tendencies. Nickel, Co and Ir and compatible ($D_{\text{SS/LS}} > 1$) in solid silicates (or oxides); W, P and Mo behave as very incompatible elements ($D_{\text{SS/LS}} < 1$); and other elements are intermediate in their silicate-system compatibility.

Physical conditions of core formation. The fractions of metal and silicate in the Earth are taken to be 0.3 and 0.7, respectively, the relative masses of the present core and mantle. The proportions of solid metal and liquid metal are determined by assuming that the principal light element in the core is sulphur. Our experiments (see below) were conducted so that our metallic liquids contained ~ 25 wt% S. Thus, if the present core contains 8–12 wt% S (ref. 18), use of our experimental data in model calculations requires that the metal assemblage which segregated to form the core was 30–50% molten at the time of metal–silicate equilibration.

The degree of partial melting of silicate materials during core formation is virtually unknown. If the Earth's accretion and core formation occurred rapidly, it seems inconceivable that high degrees of partial melting could have been avoided¹; however, the mantle samples themselves give no evidence for extensive melting early in the Earth's history. Spinel lherzolites throughout the world are very similar in composition and

texture¹⁹. If lherzolites are undepleted in elements that are strongly concentrated in basalts, these lherzolites are described as 'fertile' (that is, able to produce basalt). Large chemical differences are not observed between the fertile spinel lherzolites and the fertile garnet lherzolites¹⁹. Small differences between garnet and spinel lherzolites may exist²⁰, but the overall chemical similarity between these rock-types implies a mantle which is chemically homogeneous down to at least the depth of the sources of garnet lherzolites (~150 km; ref. 21). If large degrees of partial melting (>20%) occurred in the early Earth, even minor density differences between solid and melt would be expected to lead to phase separation and the formation of a magma ocean²². Subsequent crystallization of this hypothetical magma ocean would be unlikely to result in a chemically homogeneous upper mantle. Rather, it would be expected that the formerly molten portion of the mantle would resemble large mafic intrusions with pronounced vertical zoning, as observed in the Stillwater and Skaergaard intrusions and inferred for the Moon^{23,24}. Even if such zoning was largely homogenized by subsequent mantle convection, it is surprising that apparently no remnants remain within the continental lithospheres. The uniform chemical composition of mantle lherzolites (the samples from which our mantle siderophile and chalcophile element abundances are mainly derived), coupled with plausible physical theories of core formation which predict continuous core formation as the Earth grows from the mass of Mars to its present mass³, lead us to opt for low degrees (≤20%) of silicate partial melting.

The ambient oxygen fugacity during core formation was most likely below the iron-wüstite (IW) oxygen buffer (see later discussion), but is otherwise imprecisely constrained. Thus we allow for isothermal variation in oxygen fugacity in our core-formation model. In practice, this means that we allow $D_{LS/LM}$ to change with oxygen fugacity in an ideal manner, as defined by Rammensee²⁵. The ${}^iD_{LS/LM}$ values in Table 3 are appropriate for a log f_{O_2} of -12.75 at 1,250–1,270 °C (approximately one log unit below the QFI oxygen buffer).

Model calculations. The inefficient core formation hypothesis may be tested mathematically using the formalism outlined below:

(1) After equilibration among the four relevant phases, most of the solid and liquid metal drains away and does not communicate further with the overlying mantle. This process is subject to the constraint of mass balance, so that

$${}^iC_{Earth} = {}^iC_{SM}X_{SM} + {}^iC_{LM}X_{LM} + {}^iC_{SS}X_{SS} + {}^iC_{LS}X_{LS} \quad (1)$$

where X s are mass fractions, iC s are mass concentrations of element i and SM, LM, SS and LS refer to solid metal, liquid metal, solid silicate and liquid silicate, respectively. Reiterating the constraints outlined above,

$$X_{SS} + X_{LS} = 0.7 \quad (2)$$

$$p = X_{LS}/(X_{SS} + X_{LS}) \leq 0.2 \quad (3)$$

$$X_{SM} + X_{LM} = 0.3 \quad (4)$$

$$0.3 \leq X_{LM}/(X_{LM} + X_{SM}) \leq 0.5 \quad (5)$$

where p is the fraction of the silicate portion of the system which is molten. In terms of partition coefficients (D) and degree of partial melting, equation (1) becomes

$${}^iC_{Earth} = {}^iC_{LM}[{}^iD_{SM/LM}(0.3 - X_{LM}) + X_{LM} + 0.7{}^iD_{SS/LM}(1 - p) + 0.7{}^iD_{LS/LM}p] \quad (6)$$

As partition coefficients have been measured, ${}^iC_{Earth}/{}^iC_{LM}$ may be calculated for assumed values of p and X_{LM} . The sensitivity of this calculation to the assumptions of equations (2)–(5) should, in general, be acceptably small. For example, the permissible variation in the total masses of the various reservoirs (solid metal, liquid metal, and so forth) is probably about a factor of

two, and the effects of factor-of-two uncertainties translate into two-fold changes in calculated concentrations.

(2) A small amount of trapped solid metal (X'_{SM}) and trapped liquid metal (X'_{LM}) remains behind in the upper mantle. The concentration of element i in the mantle is then given by

$${}^iC_{mantle} = {}^iC_{LM}X'_{LM} + {}^iC_{SM}X'_{SM} + {}^iC_{SS}(1 - p) + {}^iC_{LS}p \quad (7)$$

if X'_{SM} and X'_{LM} are small (~0.01).

Rearranging as in equation (6), we obtain

$${}^iC_{mantle} = {}^iC_{LM}(X'_{LM} + {}^iD_{SM/LM}X'_{SM} + {}^iD_{SS/LM}(1 - p) + {}^iD_{LS/LM}p) \quad (8)$$

(3) As ${}^iC_{mantle}$ is known from mantle nodules and basalts, ${}^iC_{LM}$ may be calculated for assumed values of p , X'_{LM} and X'_{SM} . If ${}^iC_{LM}$ is known then ${}^iC_{Earth}$ can also be calculated from equation (6). The credibility of the chosen values for the model parameters may be evaluated by comparing the calculated ${}^iC_{Earth}$ to the concentration of element i in CI chondrites (${}^iC_{CI}$). Values of ${}^iC_{Earth}$ for refractory elements which deviate by an order of magnitude from $2-3 \times {}^iC_{CI}$ are defined as unacceptable, and new values for the model's parameters are chosen.

There are five important adjustable parameters: X_{LM} , p , f_{O_2} , X'_{LM} and X'_{SM} . However, it must be emphasized that it is unknown in detail if the measured partition coefficient values are exactly appropriate. We note that low-pressure (1 bar) partition coefficients may be relevant, in that new metal accreting to the Earth is deposited at the surface, and metal-silicate equilibrium is likely to be achieved at relatively shallow depths before the metal largely segregates into the centre of the growing planet. Nevertheless, we should probably be content if any set of model parameters can be found which, assuming an Earth with chondritic relative proportions of refractory elements ($2-3 \times CI$), will predict mantle abundances of siderophile and chalcophile elements to within a factor of two. Better fits, while mathematically pleasing, may have no physical significance.

Results. Jones and Drake⁷ have evaluated inefficient core formation as a means of reconciling the high abundances of siderophile and chalcophile elements in mantle materials with a reducing (that is, low- f_{O_2}) core-forming event. Early results were encouraging, but increasing the number of constraints on the model (that is, increasing the number of elements to be modelled) has led to progressively poorer model results. Figure 2 shows a 'best fit' calculation of bulk-Earth concentrations for our suite of siderophile and chalcophile elements. The refractory elements W to Mo have concentrations of $\sim 2 \times CI$ in the bulk Earth, a reasonable value. The more volatile elements show various degrees of depletion. Au and P concentrations appear low compared with those of more volatile elements.

Calculated Co and Ni abundances are such that the Co/Ni ratio in the Earth is within 10% of chondritic (a requirement of our model), but this agreement is somewhat contrived as Co and Ni are not concentrated in the same phase. Nickel is contained mainly in the metallic phases, whereas Co is retained by the silicates. Maintaining a chondritic Ni/Co ratio in the bulk Earth by our model requires trapping large amounts of metallic liquid (~2.5 wt %) in the mantle. Although this amount of trapped liquid is not inconsistent with theoretical estimates²⁶, the retention of this S-bearing liquid leads to unreasonably high S concentrations (~6,000 p.p.m.) in the mantle, in contrast to the observed concentration of <300 p.p.m. S (ref. 19).

Furthermore, because we allow for isothermal variation in oxygen fugacity (and, therefore, in ${}^iD_{LS/LM}$ as well), our best-fit oxygen fugacity is ~1–1.3 log units more oxidizing than the oxygen fugacity necessary for the present mantle FeO concentrations to have been established by equilibration with metal at 1 bar. This disagreement is not necessarily a problem, as an oxygen buffer such as QFI will become slightly more oxidizing with pressure (by ~1 log unit per 50 kbar), but more work is needed to define the relative effects of pressure and oxygen fugacity on siderophile and chalcophile element partitioning.

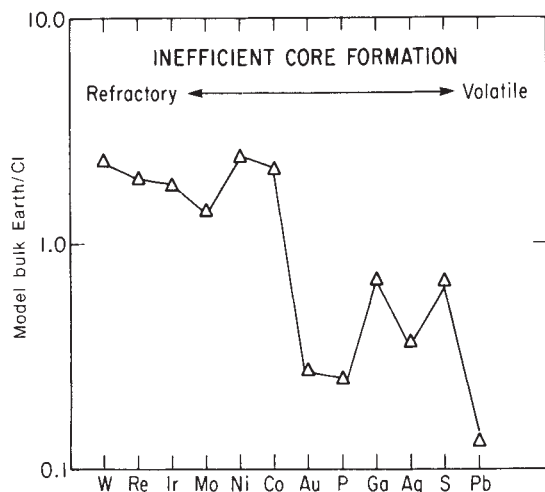


Fig. 2 Calculated bulk-Earth abundances of a suite of 12 siderophile and/or chalcophile elements. Elements are listed in the same order as in Fig. 1. The model of calculation is that of inefficient core formation (see text). The physical conditions of core formation and the amount of trapped metal phases are varied until the calculated refractory siderophile element abundances are $2\text{--}3 \times \text{CI}$ for the bulk Earth and the Ni/Co ratio of the bulk Earth is within 15% of chondritic. Clearly, geologically reasonable conditions can be chosen so that these two constraints are satisfied. A consequence of the calculation is that volatile elements are found to be depleted in the bulk Earth. A drawback of this type of model is that calculated S abundances for the upper mantle are >20 times higher than is actually observed. Best-fit model parameters: trapped metallic liquid (25 wt% S), 2.5%; trapped metal, 0.04%; degree of silicate partial melting, 10%; $\log f_{\text{O}_2} = -12.35$; S concentration in core, 10%.

Thus, even though it is possible to calculate bulk-Earth concentrations of many siderophile and chalcophile elements which are not unreasonable, the consequences of inefficient core formation are not totally attractive. If this model is to be viable, there must be substantial changes in the values of partition coefficients due to an uninvestigated intensive variable (such as pressure). Such variations are certainly not out of the question, especially as the Fe-FeS eutectic composition is known to be pressure-sensitive²⁷.

Equilibrium between an Fe-S-O metallic liquid and mantle silicates. This model has been proposed by Brett⁸, and accounts successfully for the upper-mantle abundances of a subset of siderophile and chalcophile elements. The model postulates equilibrium between metal and silicate phases and does not resort to trapping small fractions of metal in the mantle. The main difference between this model and previous equilibrium core-formation models is that no solid metal is present. It recognizes that liquid metal/silicate partition coefficients may be lower than solid metal/silicate partition coefficients (Table 3). The redox conditions postulated by Brett⁸ are rather oxidizing, in accordance with the absence of solid metal. This model is mathematically testable using partition coefficients and mass balance constraints in a similar manner to that outlined above. Generalizations concerning element partitioning are also as above.

Although at first sight this hypothesis seems successful, there are significant difficulties which appear to invalidate it. For example, the refractory element Mo was calculated by Brett⁸ to be present in the Earth's mantle at $59 \times \text{CI}$ abundance. Although this value is much closer than typical fits obtainable with solid metal/silicate partition coefficients, we consider this to be an unacceptably high value. In addition, although unknown to Brett⁸ at the time, the value of 3.65×10^{-3} which he used for the Ni liquid silicate/liquid metal partition coefficient is very large

(compare with our value of 2×10^{-4} in Table 3) and, in turn, implies an oxygen fugacity too high to be consistent with the low FeO content of the Earth's upper mantle. For example, at $1,250^\circ\text{C}$, we predict that the oxygen fugacity required for Brett's Ni partition coefficient is ~ 1 log unit more oxidizing than the Fe-FeO (IW) oxygen buffer. Under these conditions, the activity of Fe in an Fe-S-O metallic liquid will be much less than the activity of FeO in any coexisting silicates. The very low iron content of the upper mantle (<10 wt %) implies that any metallic liquid postulated to have been in equilibrium with the upper mantle (under oxidizing conditions) had an even lower iron activity than the present mantle.

The possibility that core formation in the upper mantle occurred by separation of a metallic liquid with a very low iron activity deserves further exploration. In the Fe-Si system at $1,600^\circ\text{C}$, for example, the iron activity at the midpoint of the binary system is ~ 0.05 (ref. 28). In the Fe-S system at $1,600^\circ\text{C}$ the iron activity at the midpoint of the system is ~ 0.13 (ref. 29). Non-ideal interactions between iron and non-metals can drastically reduce the activity of iron, even at fairly high iron concentrations in the liquid. Conversely, decreasing the temperature may minimize the non-ideal effects of non-metals. For example, at the eutectics of the Fe-S and Fe-Si systems, where the mole fraction of iron is definitely less than unity, iron metal is stable and the activity of iron in the liquid must also be unity. The complex interaction of intensive variables (pressure, temperature, non-metal concentrations, oxygen fugacity) make quantitative prediction of liquid activities difficult.

Although we cannot rule out highly oxidizing conditions during core formation on physical-chemical grounds, we note that fractionations of the noble refractory siderophile elements are much more probable in this model than in other models of core formation. Because noble siderophile elements appear to be present in chondritic relative abundances in the Earth's mantle, this is an important consideration. In the inefficient core formation model, mantle abundances of the noble refractory siderophile elements are 'buffered' towards chondritic ratios by trapped metal, contrary to the assertion of Morgan³⁰. In 'late-stage veneer' models, of course, no noble siderophile element fractionations are possible because the material is simply added to the upper mantle without chemical processing.

Figure 3 shows the results of recalculating bulk-Earth abundances for Brett's⁸ model by extrapolating our internally consistent set of partition coefficients to higher f_{O_2} . In this model $X_{\text{LM}} = 0.3$, the degree of silicate partial melting is 20%, and f_{O_2} is allowed to vary between IW and QFM (quartz-fayalite-magnetite, an oxidizing buffer at which iron metal is unstable). Our inefficient core formation model results from Fig. 2 are shown for comparison. The refractory siderophiles (W, Re, Ir, Mo), rather than being enriched relative to CI, are severely depleted. We conclude that Brett's⁸ equilibrium core formation model is no more viable than our own, although this conclusion is subject to the same caveat as before—that high-pressure partitioning data are virtually non-existent.

Heterogeneous accretion/'chondritic veneer'. The most recent advocates of this type of model are Morgan *et al.*⁶ and Wänke⁹, who have suggested that material accreted to the Earth became progressively more oxidized as accretion proceeded. Early accretion and core formation was dominated by highly reduced materials, and metal effectively segregated siderophile and chalcophile elements to the centre of the growing planet. By the end of accretion, an influx of more oxidized materials had caused core formation to cease because metal was no longer stable. In this model, moderately siderophile elements such as Ni and Co stopped segregating to the core when the Earth's accretion was 80–90% complete; the noble siderophile elements such as Ir and Au ended their segregation into the core when accretion was $\sim 99\%$ complete. This model predicts that moderately siderophile elements should be present in the mantle in chondritic ratios, at abundances of $0.1\text{--}0.2 \times \text{CI}$, and that the noble

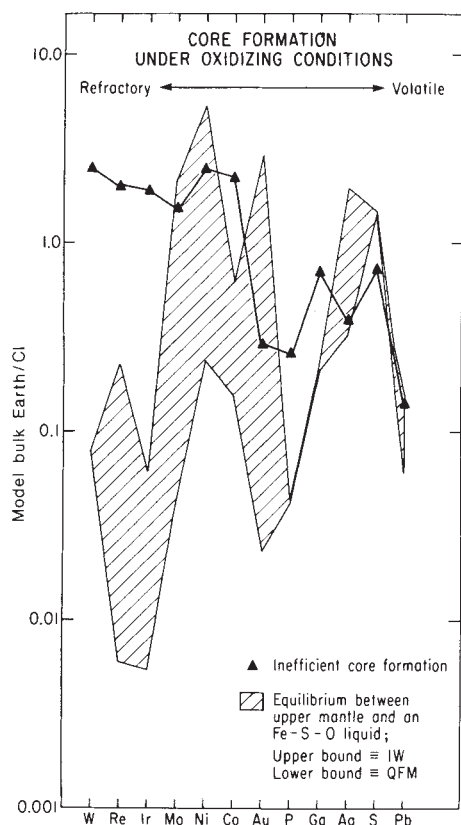


Fig. 3 Calculated bulk-Earth concentrations of siderophile and chalcophile elements assuming equilibrium between an Fe-S-O liquid and a mantle which is 20% partially molten ($T = 1,250^\circ\text{C}$). The shaded region shows the calculated concentrations for a range of oxygen fugacities, ranging from IW (upper bound) to QFM (lower bound). Bulk-Earth abundances calculated by the inefficient core formation model are shown for comparison ($\blacktriangle$). The assumption that the core formed in relatively oxidizing conditions, without any trapping of metals by the mantle, leads to the conclusion that the Earth is depleted in refractory siderophile elements, contrary to the expectations based on Fig. 1.

siderophile elements should be present in chondritic ratios at $\sim 0.01 \times \text{CI}$ abundances.

This class of hypothesis has gained favour in the past decade, in spite of the requirement that extensive physical mixing of material into the upper mantle must occur after core formation. Impact processes alone seem incapable of such fine-scale mixing (witness the heterogeneity of meteorite-derived siderophile elements in lunar samples), but sustained sub-solidus mantle convection may be adequate, particularly if mantle convection is layered. The mixing process must be capable of homogenizing Fe, Co and Ni at the scale of basalt generation by 3,500 Myr BP and be capable of homogenizing Ir at the hand-sample scale by today^{7,31}.

In this regard, it is possible that samples of the upper mantle which have been stored in continental lithospheres for considerable times (without the benefit of sub-solidus convective mixing over geological time) place even tighter constraints on the rate of mixing required to homogenize mantle siderophile elements. Jagoutz *et al.*³² have presented evidence that 2,700 Myr-old eclogites have been stored in the lower continental lithosphere for long periods of time (100–2,000 Myr) without participating in mantle mixing processes. More dramatically, Boyd *et al.*²¹ have argued that eclogitic and garnet peridotitic material, transported to the surface by kimberlite eruption, was emplaced in the lower continental lithosphere as long ago as 3,000–3,500 Myr. Thus, it is quite possible that the isolation of con-

Table 4 Test of the multi-component accretion model

Element	Concentration before second episode of core formation (C/CI)	Concentration after removal of metal* (C/CI)
Ni	0.2	0.2
Co	0.2	0.2
P	0.2	0.2
Mo	0.2	0.064
Ga	0.2	0.2
W	0.2	0.2

* Just enough metal is removed to lower the Mo concentration from 0.2 to $0.064 \times \text{CI}$ (see text).

tinental shield lithospheres from mantle convection and the homogenization of Ir that we observe today in continental lithospheric materials were both complete by 3,500 Myr ago. We question whether mantle convection was capable of such rapid, fine-scale mixing.

Another way of making the same observation is to note that Ir contents of lherzolites from beneath old cratons (such as Southern Africa), from beneath younger continental margins (such as the southwestern United States) and from mantle overlying oceanic hot-spots (Hawaii) are essentially identical¹⁹. The lithospheres beneath southern Africa and the southwestern United States were not obviously isolated from mantle convection at the same time, and the lithosphere beneath Hawaii has presumably never been isolated to the same degree as continental lithospheres. Iridium homogenization was apparently complete by the time of the formation of the earliest continental lithospheres.

Regardless of absolute rates of mantle mixing, the moderately siderophile elements are a sensitive test of the heterogeneous accretion hypothesis. Table 4 shows our model results using the partition coefficients of Table 3. For simplicity we have quantized accretion and core formation (before the accretion of the last 1%—the hypothesized 'late-stage veneer') into two parts: the first 80% of material accreted is highly reduced, the last 20% is more oxidized. This approach permits us to model core formation as two discrete events: the first, which removed all siderophile and chalcophile elements from the early-accreted 80% of the Earth, and a second event which removed only noble siderophile elements (such as Ir and Au) from the late-accreted 20%. A justification of this simple approach is that, because of the large difference in the metal/silicate partition coefficients between Co and Ni, continuous core formation during the last 20% of accretion would be unlikely to preserve the approximately chondritic Co/Ni ratio of the upper mantle (Table 2). Thus, at the beginning of the second model core-forming event, most (if not all) siderophile and chalcophile elements were present at $\sim 0.2 \times \text{CI}$ abundances in the upper mantle, as defined by present Ni and Co abundances. At this time we allow just enough separation of metal to lower the Mo concentration from $0.2 \times \text{CI}$ to its present value of $0.064 \times \text{CI}$ (Table 2), and we observe the effect of this separation on the other siderophile elements. As can be seen from Table 4, no moderately siderophile element other than Mo is affected, even though W and P are clearly depleted relative to Co and Ni in the present mantle (Table 2).

The heterogeneous accretion hypothesis results in additional predictions which are inconsistent with observed elemental abundances in the Earth's upper mantle. First, the amount of metal which was extracted from the last 20% of accreting material must have been trivial, and the high f_{O_2} requires that that metal must have been Ni-rich (because of the decreased stability of Fe-rich metal). If we assume that as much as 10% of the total Ni was removed during the last core-forming event (consistent with the approximately chondritic Co/Ni ratio of the upper mantle), then only 0.02% metal was removed. It would

seem that a substantially molten upper mantle would be required to separate this minute amount of metal, and, as we have already discussed, there is no evidence for a pervasive terrestrial magma ocean. Second, if the last 20% of accreted material was sufficiently oxidized to effectively halt core formation, it is expected that this material would have been enriched in volatile elements and oxidized species. If the last dregs of core formation removed only a few hundredths of a per cent of metal, it is not possible for substantial amounts of S to have been taken to the core (S is cosmically more abundant than Ni). All chondritic materials contain 2–6 wt% S. If the material which made up the last 20% of the Earth's accretion even approached chondritic composition, we would expect that the crust and upper mantle would contain 4,000–12,000 p.p.m. S: not appreciably different from the prediction of the inefficient core formation hypothesis, and in contradiction to observed S abundances of <300 p.p.m. in the Earth's upper mantle¹⁹.

In this context we note that Morgan³³ has refined his abundance estimates of incompatible noble siderophile elements (for example, Au and Re) in spinel lherzolites, and finds that the noble element ratios are remarkably chondritic. Obviously, this observation strengthens the arguments for a 1% chondritic 'late-stage veneer'. Furthermore, Morgan shows that the S/Se ratio of spinel lherzolites is chondritic and advocates that the S and Se abundances of the upper mantle were established by the same 'late-stage veneer' process. The implication of Morgan's work appears to be that, before the last 1% of accretion, the mantle was sulphur-free. In principle, this may remove some of the difficulties with S which were discussed above, although it is still unclear whether such low S abundances during the last 20% of the Earth's inhomogeneous accretion are compatible with the more oxidizing conditions postulated for this stage of accretion⁹.

We note that the consistency of the siderophile element abundances inferred from basalts, spinel lherzolites and garnet lherzolites implies that the siderophile elements of the upper mantle are likely to be homogeneously distributed to a depth of at least 500 km—a scale consistent with models of upper mantle convection³⁴. If 1% of the mass of the upper mantle were added by a uniform flux of planetesimals and if this same population were responsible for the production of a 2.5-km megaregolith on the Moon²⁴, we would expect that the lunar megaregolith would contain ~200% of chondritic material—a physically impossible value which is about two orders of magnitude higher than the ~1.5 wt% of meteoritic material measured in lunar soils²⁴. The Moon would have to be well-mixed to a depth of 400 km in order to be consistent with the terrestrial 'late-stage veneer', yet estimates of mantle abundances of noble siderophile elements¹³ show no evidence of such mixing. Of course, it is possible that the Moon formed after the Earth accreted its final 1% of mass, although available isotopic evidence points to comparable ages for both objects³⁵.

Although we find some elements of the heterogeneous accretion/'chondritic veneer' hypothesis attractive, this hypothesis contains as many quantitative difficulties as the inefficient core formation and metallic liquid/silicate equilibrium core formation hypotheses. It is not clear to us which, if any model of core formation should be preferred.

Conclusions

The preceding discussion emphasizes that none of the current hypotheses of core formation in the Earth survives quantitative scrutiny. For any of the hypotheses to be correct, our present knowledge of the relevant relative partition coefficients must be accurate to perhaps no better than a factor of two. We emphasize that the volume of ($P-T-f_{O_2}$) space which has been investigated in our partitioning studies (as well as those of others) is small and should be extended.

This failure of all hypotheses to quantitatively model core formation and siderophile and chalcophile element abundances

in the upper mantle is disappointing, in that a detailed picture of the history of the early Earth is important to our understanding of many consequential issues, such as the origin of the Moon and the subsequent geological evolution of the Earth. What, if any, statements about core formation may be made with such confidence that they will go unchallenged?

Lack of equilibration between core and upper mantle. All viable models for the explanation of the high abundances of siderophile and chalcophile elements in the mantle implicitly assume that the core and upper mantle have not communicated over geological time. An immediate consequence of this assumption is that either convection in the mantle is layered and the upper and lower mantles communicate only with great difficulty, or the mantle which we sample today was incorporated into the subcontinental lithosphere extremely early. In fact, these two possibilities are not mutually exclusive; in any case the upper mantle and core have not equilibrated since the last addition of siderophile elements to the upper mantle.

Antiquity of siderophile and chalcophile elements in the upper mantle. However the abundances of siderophile and chalcophile elements were established, they are unlikely to have changed over geological time. Evidence from basalts indicates that the Co, Fe and Ni contents of the upper mantle have not changed since 3,500 Myr (ref. 31) and that 'future leads' cannot be explained in terms of continuous core formation³⁶. Siderophile and chalcophile element concentrations in the upper mantle were established early in the history of the Earth and are not the result of secondary processes.

The redox state of the early Earth. All models of core formation seem to require that the mantle has become more oxidized over geological time. If the FeO content of the upper mantle is the result of equilibration with metal, then the consistent FeO content of 8–10 wt% in fertile mantle nodules appears to limit the oxidation state of the early Earth to oxygen fugacities less than ~1.5–2.0 log units below the QFI buffer. (For a discussion of the assumptions involved in this calculation see ref. 37.) This estimate should be accurate and is tantalizingly close to the oxygen fugacity of the metal-olivine-pyroxene reaction of ref. 38. If the FeO content of the mantle was not established by reaction with metal but, rather, by later admixture of FeO-rich material, then the calculated oxygen fugacity is an upper limit to that obtaining during core formation. If the FeO content of the mantle was due primarily to reaction with a metallic liquid, rich in non-metals⁸, the oxygen fugacity was probably still within two log units of IW. In all these cases the mantle was once very reduced and its oxygen fugacity must have changed over geological time to its present oxidized state (that is, from near IW (or lower) to QFM; ref. 39). If the oxygen fugacity calculated at the beginning of this section is accurate, no reactions other than the Fe metal-olivine-pyroxene reaction³⁸ need be postulated; if the calculation is only an upper limit, it may be necessary to consider seriously the hypothesis that the Earth is predominantly made of enstatite-chondrite-like material.

Also, we note that the moderately low non-metal concentrations (8–12%) inferred for the core (compared with the high concentrations (~30%) of S and O in the metallic liquid of Brett⁸) appear to require that models of core formation which postulate large amounts of Fe-S-O-rich liquids cannot apply throughout the core formation process. If non-metal-rich metallic liquids contributed significantly to core formation, then additional metal, containing little or no S or O, is required to account for the observed non-metal concentrations in the present core. The easiest means of achieving such a dichotomy is probably through inhomogeneous accretion, in which earlier core-forming events were S-O-poor and later events were S-O-rich. Thus, even models which postulate oxidizing conditions for core formation may require that the oxygen fugacity changed during planetary accretion and core formation, becoming more oxidizing with time.

Future directions. We emphasize again that only a very small

volume of (P, T, f_{O_2})-space relevant to core formation has been investigated. Two obvious areas for new experimental explorations are those of high pressure and higher oxygen fugacity. For example, from our work on iron meteorites⁴⁰, we have gained some insight into the possible role of pressure. Many siderophile trace element-non-metal interactions in metallic liquids appear to be explicable in terms of a simple non-metal-avoidance model. Siderophile trace element-sulphur interactions are often ~10 times stronger than Fe-S interactions, which are known to be pressure-sensitive. (Specifically, the composition of the eutectic point in the Fe-FeS binary system becomes more Fe-rich as

pressure increases²⁷.) Extrapolating from the Fe-S system, it is possible that siderophile element-non-metal interactions will be strong functions of pressure. Quantitatively successful models of core formation may require accurate and detailed assessments of the (P, T) state of the early Earth.

This work was supported by NASA grant NAG 9-39. We thank H. Doane and the Reactor Facility of the University of Arizona for assistance; D. Hill and W. V. Boynton for help with neutron activation analyses; D. Stevenson and F. Richter for discussions; and R. Brett, D. J. Malvin, J. W. Morgan and H. E. Newsom for helpful reviews.

1. Brett, R. *Rev. Geophys. Space Phys.* **14**, 375-383 (1976).
2. Urey, H. C. *The Planets* (Yale University Press, New Haven, 1952).
3. Stevenson, D. J. *Science* **214**, 611-619 (1981).
4. Ringwood, A. E. *Origin of the Earth and Moon* (Springer, Berlin, 1979).
5. Ringwood, A. E. *Geochim. cosmochim. Acta* **35**, 223-230 (1971).
6. Morgan, J. W., Wandless, G. A., Petrie, R. K. & Irving, A. J. *Proc. Lunar planet. Sci. Conf.* **11**, 213-233 (1980).
7. Jones, J. H. & Drake, M. J. *Eos* **62**, 1073-1074 (1981); *Proc. Lunar planet. Sci. Conf.* **13**, 369-370 (1982); **15**, 413-414 (1984); **16**, 412-413 (1985).
8. Brett, R. *Geochim. cosmochim. Acta* **48**, 1183-1188 (1984).
9. Wänke, H. *Phil. Trans. R. Soc. A* **303**, 287-302 (1981).
10. Jones, J. H. & Drake, M. J. *Geochim. cosmochim. Acta* **47**, 1199-1209 (1983).
11. Jagoutz, E. et al. *Proc. Lunar planet. Sci. Conf.* **10**, 2031-2050 (1979).
12. Chou, C.-L., Shaw, D. M. & Crockett, J. H. *J. geophys. Res.* **88**, A507-A518 (1983).
13. Drake, M. J. *Geochim. cosmochim. Acta* **47**, 1759-1767 (1983).
14. Newsom, H. E. & Drake, M. J. *Geochim. cosmochim. Acta* **47**, 93-100 (1983).
15. Newsom, H. E. & Palme, H. *Earth planet. Sci. Lett.* **69**, 354-364 (1984).
16. Laul, J. C., Keays, R. R., Ganapathy, R., Anders, E. & Morgan, J. W. *Geochim. cosmochim. Acta* **36**, 329-345 (1972).
17. Hertogen, J., Janssens, M.-J. & Palme, H. *Geochim. cosmochim. Acta* **44**, 2125-2143 (1980).
18. Brown, J. M., Ahrens, T. J. & Shampine, D. L. *J. geophys. Res.* **89**, 6041-6048 (1984).
19. Basaltic Volcanism Study Project *Basaltic Volcanism on the Terrestrial Planets* (Pergamon, Oxford, 1981).
20. Palme, H. & Nickel, K. G. *Geochim. cosmochim. Acta* **49**, 2123-2132 (1985).
21. Boyd, F. R., Gurney, J. J. & Richardson, S. H. *Nature* **315**, 387-389 (1985).
22. Walker, D., Stolper, E. M. & Hays, J. F. *J. geophys. Res.* **83**, 6005-6013 (1978).
23. Hofmeister, A. M. *J. geophys. Res.* **88**, 4963-4983 (1983).
24. Taylor, S. R. *Lunar Science: A Post-Apollo View* (Pergamon, Oxford, 1975).
25. Rammensee, W. thesis. Univ. Mainz (1978).
26. Stevenson, D. J. in *Workshop on the Early Earth: The Interval from Accretion to the Older Archean* (eds Burke, K. & Ashwal, L. D.) 76-78 (LPI tech. Rep. 85-01, Lunar and Planetary Institute, Houston, 1985).
27. Usselman, T. M. *Am. J. Sci.* **275**, 278-290 (1975).
28. Hultgren, R., Prasad, D. D., Hawkins, D. T., Gleiser, M. & Kelley, K. K. *Selected Values of the Thermodynamic Properties of Binary Alloys*, 871-883 (American Society of Metals, Metals Park, 1973).
29. Sharma, R. C. & Chang, Y. A. *Mettall. Trans.* **10B**, 103-108 (1979).
30. Morgan, J. W. *Nature* **317**, 703-705 (1985).
31. Delano, J. W. & Stone, K. *Proc. Lunar planet. Sci. Conf.* **16**, 181-182 (1985).
32. Jagoutz, E., Dawson, J. B., Hoernes, S., Spettel, B. & Wänke, H. in *Workshop on the Early Earth: The Interval from Accretion to the Older Archean* (eds Burke, K. & Ashwal, L. D.) 40-41 (LPI tech. Rep. 85-01, Lunar and Planetary Institute, Houston, 1985).
33. Morgan, J. W. *J. geophys. Res.* (in the press).
34. Richter, F. M. *Earth planet. Sci. Lett.* **73**, 350-360 (1985).
35. Swindle, T. D., Caffee, M. W., Hohenberg, C. M. & Taylor, S. R. in *Origin of the Moon* (Lunar and Planetary Institute, Houston, in the press).
36. Newsom, H. E., White, W. M. & Jochum, K. P. *Proc. Lunar planet. Sci. Conf.* **16**, 616-617 (1985).
37. Newsom, H. E. *J. geophys. Res.* **90**, C613-C617 (1985).
38. Larimer, J. W. *Geochim. cosmochim. Acta* **32**, 1189-1209 (1968).
39. Arculus, R. J. A. *Rev. Earth planet. Sci.* **13**, (1985).
40. Malvin, D. J., Jones, J. H. & Drake, M. J. *Geochim. cosmochim. Acta* (in the press).
41. Meyers, J. & Eugster, H. P. *Contr. Miner. Petrol.* **82**, 75-90 (1983).
42. Zartman, R. & Tera, F. *Earth planet. Sci. Lett.* **20**, 54-66 (1973).
43. Anders, E. & Ebihara, M. *Geochim. cosmochim. Acta* **46**, 2363-2380 (1982).
44. Irving, A. J. *Geochim. cosmochim. Acta* **42**, 743-770 (1978).
45. Hertogen, J., Janssens, M.-J., Takahashi, H., Morgan, J. W. & Anders, E. *Geochim. cosmochim. Acta* **47**, 2241-2255 (1983).
46. Sears, D. W., Kallemeyn, G. W. & Wasson, J. T. *Geochim. cosmochim. Acta* **46**, 597-608 (1982).
47. Von Michaelis, H., Willis, J. P., Erlank, A. J. & Ahrens, L. H. *Earth planet. Sci. Lett.* **5**, 383-386 (1969).
48. Moore, C. B. in *Handbook of Elemental Abundances in Meteorites* (ed. Mason, B.) 137-142 (Gordon and Breach, New York, 1971).
49. Manhès, G. & Allègre, C. J. *Meteoritics* **13**, 543-548 (1978).
50. Kallemeyn, G. W. & Wasson, J. T. *Geochim. cosmochim. Acta* **45**, 1217-1230 (1981).
51. Anders, E., Higuchi, H., Ganapathy, R. & Morgan, J. W. *Geochim. cosmochim. Acta* **40**, 1131-1139 (1976).
52. Morgan, J. W. et al. *Nature* **224**, 789-791 (1969).
53. Palme, H. & Rammensee, W. *Earth planet. Sci. Lett.* **55**, 356-362 (1981).
54. Mason, B. *Space Sci. Rev.* **1**, 621-646 (1963).
55. Tatsumoto, M., Unruh, D. M. & Desborough, G. A. *Geochim. cosmochim. Acta* **40**, 617-634 (1976).
56. Mason, B. *Handbook of Elemental Abundances in Meteorites*, **55**, (Gordon and Breach, New York, 1971).
57. Laul, J. C., Ganapathy, R., Anders, E. & Morgan, J. W. *Geochim. cosmochim. Acta* **37**, 329-357 (1973).
58. Tatsumoto, M., Knight, R. J. & Allègre, C. J. *Science* **180**, 1279-1283 (1973).

ARTICLES

Binding of a specific ligand inhibits import of a purified precursor protein into mitochondria

Martin Eilers & Gottfried Schatz

Biocenter, University of Basel, CH-4056 Basel, Switzerland

Methotrexate, a folate antagonist, blocks import into mitochondria of mouse dihydrofolate reductase fused to a mitochondrial presequence. Methotrexate does not mask the presequence, but stabilizes the dihydrofolate reductase moiety. It does not inhibit import of the authentic precursor from which the presequence is derived. This suggests that dihydrofolate reductase must at least partly unfold in order to be transported across mitochondrial membranes.

TRANSLOCATION of proteins across biological membranes is a key step in the intracellular sorting and secretion of proteins. Although much is known about protein movement across various membrane systems¹, it is unclear how hydrophilic, charged proteins translocate through the hydrophobic core of a phos-

pholipid bilayer. In particular, it remains unanswered whether a protein must unfold during the transport process² or whether it can translocate in a folded state³, perhaps through transient non-bilayer domains in the target membrane⁴. We have approached this question by fusing the presequence of an im-

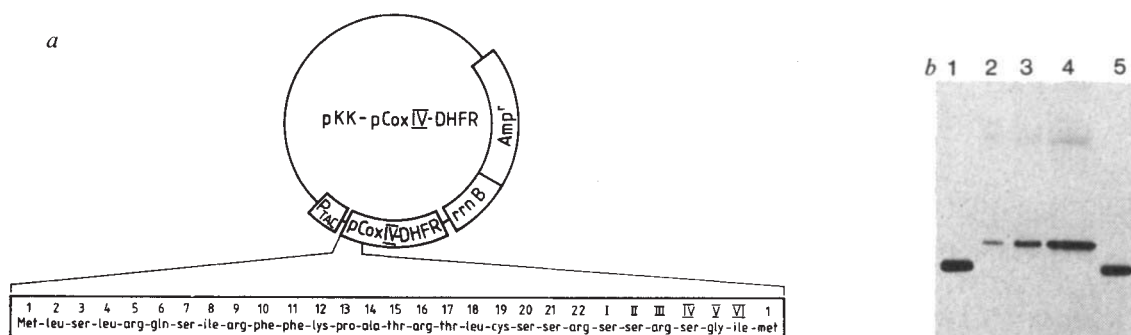


Fig. 1 Expression and purification of mouse DHFR carrying a mitochondrial presequence. The fusion gene encoding the first 22 residues of the yeast cytochrome oxidase subunit IV presequence linked to mouse DHFR⁵ was excised from the yeast-*E. coli* shuttle vector pLGS-D5-pCoxIV-DHFR⁶ with *Bam*HI and treated with *Bal*31 to remove flanking sequences. The ends were filled-in with the large (Klenow) fragment of *E. coli* DNA polymerase and the shortened fusion gene was cut with *Hind*III at its 3' end. A fragment of the expected size was isolated²⁰ and cloned into a 'TAC'-expression vector²¹ (pKK 223-3; Pharmacia) using a filled-in *Eco*RI site at one end and a *Hind*III site at the other. *E. coli* HB101 (ref. 22) transformants resistant to antifolates (500 μ g trimethoprim ml⁻¹ or 5–10 μ g 1,3-diamino-7-benzylpyrrolo-[3,2/*f*]-quinazoline ml⁻¹) were screened by colony immunoblotting²³ using a polyclonal antiserum against mouse DHFR. Plasmid pKK-pCoxIV-DHFR (a) was isolated from a transformant giving a strong signal with the anti-mouse DHFR antibody. P_{TAC}, TAC-promoter; rrrB, ribosomal transcription terminator; Amp^r, β -lactamase gene. The insert shows the first 29 amino acids of the fusion protein, consisting of 22 amino acids of the subunit IV presequence (1–22), 6 amino acids introduced by the cloning procedure (I–IV) and the first methionine of mouse DHFR. *E. coli* transformed with plasmid pKK-pCoxIV-DHFR were grown to the late logarithmic phase on LB medium²⁴, induced for 2 h at 37 °C with 1 mM isopropyl- β -thiogalactoside, and harvested. In these induced cells, the fusion protein represented ~1% of the total cellular protein. All subsequent steps were performed at 0–5 °C. The cells were osmotically shocked to release most of the periplasmic contents²⁵, then suspended in 250 mM KCl, 20 mM KP_i, pH 8.0, 1 mM EDTA, 1 mM EGTA and 0.1 mM phenylmethylsulphonyl fluoride (PMSF). The shocked cells were gently lysed by adding lysozyme to 0.4 mg ml⁻¹ and, after 30 min, octyl-polyoxyethylene (given by Dr J. Rosenbusch, Biocenter) to 1%. The lysate was diluted 10-fold with 20 mM KP_i, pH 7.0, 1 mM EDTA, 0.1 mM PMSF, mixed with 0.1 vol. of 2% protamine sulphate, stirred for 15 min, and centrifuged at 6,000g for 10 min. The supernatant was chromatographed on a CM-cellulose column in the presence of 0.5% octyl-polyoxyethylene, 0.1 mM PMSF, using a linear KCl gradient (25–500 mM). Fractions exhibiting DHFR activity²⁶ were pooled, dialysed overnight against 20 mM KP_i, pH 5.8, and subjected to affinity chromatography on methotrexate-Sepharose⁸, using 4.5 mM dihydrofolate in 1 M KCl, 20 mM KP_i, pH 8.0, as eluent. b, A silver-stained²⁷ SDS/12% polyacrylamide gel. Lanes 1 and 5, authentic mouse DHFR; 2–4, purified fusion protein (36, 108 and 360 ng, respectively). Protein was assayed as described elsewhere⁵.

ported mitochondrial protein to mouse dihydrofolate reductase (DHFR)⁵, and studying the translocation of the fusion protein into isolated yeast mitochondria.

The fusion protein

Hurt *et al.*^{5,6} have fused the first 22 residues of the presequence of yeast cytochrome oxidase subunit IV (an imported mitochondrial protein) to the N-terminus of mouse DHFR (a cytosolic protein), and shown that the resulting fusion protein is imported and proteolytically processed by isolated yeast mitochondria or by mitochondria in living yeast cells. DHFR lacking an attached presequence is not imported. In all properties studied, import of the fusion protein into mitochondria resembled import of authentic mitochondrial precursor proteins.

When we expressed the fusion protein in *Escherichia coli* (Fig. 1a), the bacterial cells became resistant to high levels of the folate antagonists trimethoprim and 1,3-diamino-7-benzylpyrrolo-[3,2/*f*]-quinazoline, indicative of elevated intracellular levels of DHFR⁷. When the cells were rapidly lysed in NaOH and analysed by immune blotting with antiserum against mouse DHFR, only undegraded fusion protein was detected (not shown); thus, this protein is enzymatically active *in vivo*.

The fusion protein was purified from the *E. coli* transformants by cation-exchange chromatography in the presence of a non-ionic detergent, followed by affinity chromatography on Sepharose derivatized with the folate antagonist methotrexate⁸. The purified fusion protein was nearly homogeneous as judged by SDS-polyacrylamide gel electrophoresis (Fig. 1b); by this procedure, 1 litre of an *E. coli* culture harvested in late logarithmic phase yielded 180 μ g of purified fusion protein.

The fusion protein appeared to aggregate within the *E. coli* cells and required detergent for efficient extraction (not shown). Once extracted, however, it remained soluble even on removal of the detergent. The purified soluble fusion protein exhibited

DHFR activity; its specific activity (13 U mg⁻¹) was lower than that of authentic mouse DHFR (31 U mg⁻¹) but the fusion protein might have retained traces of methotrexate that had leaked off the affinity column. The *K_m* values for NADPH were 0.58 μ M for the fusion protein and 0.54 μ M for authentic DHFR. For technical reasons, *K_m* values for dihydrofolate could not be measured precisely, but they, too, were similar (0.13–0.31 μ M).

We conclude that the attachment of a presequence to the amino-terminus of DHFR does not significantly alter the protein's enzymatic properties, and therefore probably does not greatly alter the protein's conformation: the presequence and attached 'mature' sequence probably represent separate domains. The experiments reported below support this view.

Blockade of fusion protein import

The purified fusion protein was imported by energized yeast mitochondria (Fig. 2a). When 35 ng of purified fusion protein was incubated with 120 μ g of de-energized yeast mitochondria, 7 ng (20%) was found in association with the mitochondria (Fig. 2a, lane 2) but was not imported: it remained accessible to externally added proteinase K (lane 3). On incubation with energized mitochondria, however, 42% (15 ng protein) of the fusion protein molecules were imported: they were cleaved to a smaller species (Fig. 2a, lane 4) which was inaccessible to externally added protease (lane 5). When the mitochondrial membranes were disrupted with detergent, the imported molecules, too, became accessible to added protease; under these conditions, they were cleaved to a proteolytic fragment (Fig. 2a, lane 6) distinctly smaller than that generated on import into mitochondria (compare lanes 5 and 6). Methotrexate (55 nM) inhibited mitochondrial import of the fusion protein by 89%; binding to the mitochondrial surface was decreased by only 17% (Fig. 2b; all lanes correspond to those of Fig. 2a

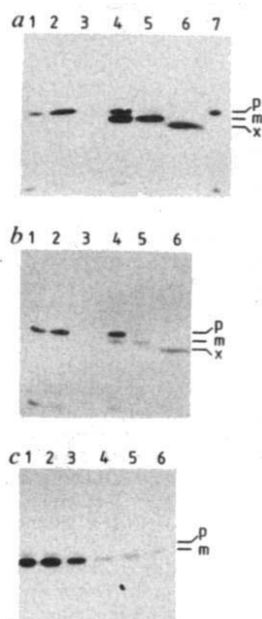


Fig. 2 Methotrexate blocks import of the fusion protein into isolated mitochondria. *a*, Lanes 1 and 7, 3.5 ng fusion protein (10% of the amount added to mitochondria); 2 and 3, fusion protein incubated with de-energized mitochondria ($1 \mu\text{g}$ valinomycin ml^{-1} was present); 4–6, fusion protein incubated with energized mitochondria. The samples shown in lanes 3 and 5 had been treated with proteinase K after import; that in lane 6 had been treated with proteinase K in the presence of 1% Triton X-100. *p* and *m*, uncleaved and cleaved fusion protein, respectively; *x*, protease-resistant degradation product of DHFR. Relative molecular masses (M_r s) were 25,000 (25K), 22K and 21K for *p*, *m* and *x*, respectively. *b*, Same as *a*, except that all incubations contained 55 nM methotrexate, and lane 7 was omitted. *c*, Import of the fusion protein into energized mitochondria in the presence of the following concentrations (nM) of methotrexate: lane 1, 0.83; 2, 2.75; 3, 8.3; 4, 27.5; 5, 83; 6, 275. Import conditions as in *a*, lane 5.

Methods. *E. coli* transformed with pKK-pCoxIV-DHFR were pulse-labelled for 30 min at 37°C with 2 mCi of carrier-free $^{35}\text{S}\text{O}_4^{2-}$ ($>1,000 \text{ Ci mmol}^{-1}$) in synthetic minimal medium supplemented with 1 mM isopropyl- β -D-thiogalactoside and all amino acids except cysteine and methionine²⁸. The fusion protein was then isolated as described in Fig. 1 legend except that affinity chromatography was omitted. Fractions eluted from CM-cellulose were analysed for fusion proteins by SDS-polyacrylamide gel electrophoresis and fluorography. Appropriate fractions were pooled and dialysed overnight against 100 mM KCl, 20 mM HEPES-KOH pH 7.4, 1 mM EDTA, 1 mM dithiothreitol, 0.1 mM PMSF. Aliquots of the purified fusion protein (35 ng; 1.5×10^9 c.p.m. per mg protein) were incubated with isolated yeast mitochondria ($120 \mu\text{g}$ protein) for 30 min at 30°C and the mitochondria were analysed for imported fusion protein²⁹. Internalization was ascertained by treating the re-isolated mitochondria with $50 \mu\text{g}$ proteinase K ($250 \mu\text{g ml}^{-1}$) for 30 min at 0°C before electrophoretic analysis. The amount of added fusion protein was quantitative by calibrated immune blotting⁶, using authentic mouse DHFR as a standard.

except that methotrexate was added to all incubations). Methotrexate inhibited import of the fusion protein at roughly the same concentration at which it inhibited the enzymatic activity of pure mouse DHFR (Fig. 2c), the K_m values for inhibition of import and of enzymatic activity being 4.1 and 3.8 nM, respectively (data not shown). The result suggests that methotrexate blocks mitochondrial import of the fusion protein by binding to the catalytic site⁹ on DHFR.

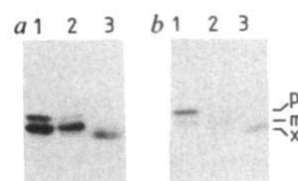


Fig. 3 A hydrophobic folate antagonist also blocks import of the fusion protein into yeast mitochondria. All incubations were carried out as described in Fig. 2 legend. *a*, Lanes 1, 2 and 3 show, respectively, fusion protein incubated with energized mitochondria and left untreated, treated with proteinase K or treated with proteinase K in the presence of 1% Triton X-100. *b*, Same as *a*, except that all incubations contained $1 \mu\text{M}$ 1,3-diamino-7-benzylpyrrolo-[3,2/*f*]-quinazoline.

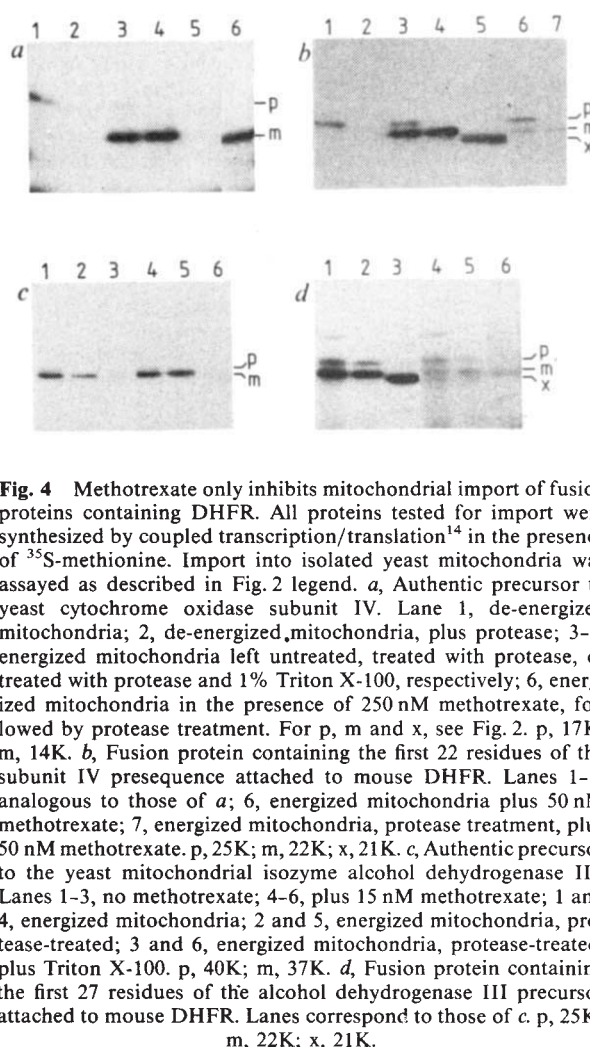


Fig. 4 Methotrexate only inhibits mitochondrial import of fusion proteins containing DHFR. All proteins tested for import were synthesized by coupled transcription/translation¹⁴ in the presence of ^{35}S -methionine. Import into isolated yeast mitochondria was assayed as described in Fig. 2 legend. *a*, Authentic precursor to yeast cytochrome oxidase subunit IV. Lane 1, de-energized mitochondria; 2, de-energized mitochondria, plus protease; 3–5, energized mitochondria left untreated, treated with protease, or treated with protease and 1% Triton X-100, respectively; 6, energized mitochondria in the presence of 250 nM methotrexate, followed by protease treatment. For *p*, *m* and *x*, see Fig. 2. *p*, 17K; *m*, 14K. *b*, Fusion protein containing the first 22 residues of the subunit IV presequence attached to mouse DHFR. Lanes 1–5, analogous to those of *a*; 6, energized mitochondria plus 50 nM methotrexate; 7, energized mitochondria, protease treatment, plus 50 nM methotrexate. *p*, 25K; *m*, 22K; *x*, 21K. *c*, Authentic precursor to the yeast mitochondrial isozyme alcohol dehydrogenase III. Lanes 1–3, no methotrexate; 4–6, plus 15 nM methotrexate; 1 and 4, energized mitochondria; 2 and 5, energized mitochondria, protease-treated; 3 and 6, energized mitochondria, protease-treated, plus Triton X-100. *p*, 40K; *m*, 37K. *d*, Fusion protein containing the first 27 residues of the alcohol dehydrogenase III precursor attached to mouse DHFR. Lanes correspond to those of *c*. *p*, 25K; *m*, 22K; *x*, 21K.

Import of the purified fusion protein into mitochondria is also blocked by 1,3-diamino-7-benzylpyrrolo-[3,2/*f*]-quinazoline, a membrane-permeant antifolate (ref. 10 and R.L. Then, personal communication) (Fig. 3). The inhibition of import of the fusion protein is, therefore, not simply a consequence of the inability of methotrexate to cross biological membranes.

To demonstrate that methotrexate is not a general inhibitor of mitochondrial protein import, we tested its effect on

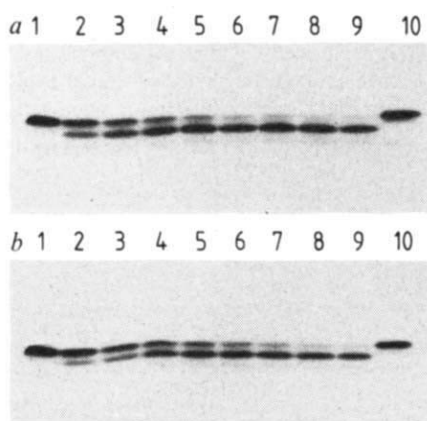


Fig. 5 Methotrexate does not inhibit the action of mitochondrial processing protease on the presequence attached to DHFR. *a*, Lane 1, no incubation with protease; lanes 2–9, incubation with protease for 1.5, 2.5, 5, 7.5, 10, 15, 20 and 30 min, respectively; lane 10, incubation for 30 min in the presence of 1 mM 1,10-phenanthroline, a relatively specific inhibitor of the matrix protease. *b*, Same as *a*, except that all incubations contained, in addition, 70 nM methotrexate.

Methods. The fusion protein containing the first 22 residues of the subunit IV presequence attached to mouse DHFR was synthesized *in vitro* in the presence of ^{35}S -methionine (see Fig. 4 legend) and incubated with the partially purified¹⁵ processing protease. Samples were precipitated with 5% trichloroacetic acid and analysed by SDS/12% polyacrylamide gel electrophoresis and fluorography.

mitochondrial import of four different proteins: (1) the authentic precursor to cytochrome oxidase subunit IV¹¹; (2) the authentic precursor to yeast alcohol dehydrogenase III (a mitochondrial isozyme¹²); (3) mouse DHFR linked to the first 22 residues of the cytochrome oxidase subunit IV presequence⁵ (that is, the fusion protein described above); and (4) mouse DHFR linked to the first 27 residues of the alcohol dehydrogenase III precursor¹³. Each of the four proteins was synthesized *in vitro* by coupled transcription-translation in the presence of ^{35}S -methionine and tested for cleavage and import by isolated yeast mitochondria. Methotrexate blocked import of only the two fusion proteins (Fig. 4*b, d*); it did not measurably affect import of the two authentic precursors whose presequences had been used to construct the fusion proteins (Fig. 4*a, c*). Thus, methotrexate blocked mitochondrial import of only those proteins containing a DHFR moiety.

Methotrexate binding

Binding of methotrexate to the fusion proteins could trigger an interaction between the DHFR domain and the attached presequence, rendering that presequence inaccessible to the mitochondrial import machinery. Two observations make this possibility unlikely. First, methotrexate did not significantly affect binding of the fusion proteins to de-energized mitochondria (Figs 2, 4); as DHFR lacking a presequence does not bind to mitochondria¹⁴, binding of the fusion proteins shows that the presequence is accessible. Second, methotrexate did not significantly inhibit removal of the attached presequence by the processing protease partially purified¹⁵ from the yeast mitochondrial matrix (Fig. 5). Methotrexate did, however, partially protect the DHFR moiety against degradation by low concentrations of thermolysin (Fig. 6). Incubation of the purified fusion protein with thermolysin in the presence of methotrexate caused accumulation of a degradation intermediate whose size corresponded to that of DHFR lacking a presequence (Fig. 6, lanes 5–11). Methotrexate also protected authentic DHFR against thermolysin; in that case, however, no smaller intermediate was detected (not shown). Thus, it is likely that the cleaved-off



Fig. 6 Methotrexate protects DHFR against degradation by thermolysin. ^{35}S -labelled purified fusion protein (7.5 ng; see Fig. 1 legend) was incubated for 30 min at 43 °C in 0.1 M HEPES-KOH pH 7.4, 2 mM CaCl_2 in the absence of protease (lane 1) or with 5.6 ng (lanes 2, 5, 8, 11), 11.2 ng (lanes 3, 6, 9, 12) or 28 ng (lanes 4, 7, 10) of thermolysin, either in the absence of methotrexate (lanes 1–4), or in the presence of 16.5 μM (lanes 5–7), 660 nM (lanes 8–10) or 66 nM (lanes 11, 12) methotrexate. The final volume in each case was 50 μl . Samples were then mixed with 5 μl of 0.2 M EGTA and analysed by SDS/12% polyacrylamide gel electrophoresis and fluorography.

fragment which is not protected from thermolysin by methotrexate is indeed the presequence. Control experiments (not shown) established that methotrexate at a concentration as high as 7 μM did not affect the action of thermolysin on yeast hexokinase or chicken ovalbumin.

Implications for protein translocation

How does methotrexate block import of the DHFR-containing fusion proteins into mitochondria? We can exclude the possibility that it blocks mitochondrial protein import *per se* or causes masking of the presequence. Almost certainly, it also does not induce a major structural change in the DHFR moiety: extensive X-ray crystallographic studies on the interaction between chicken DHFR and antifolates have revealed only a minor displacement of Tyr 30 and Glu 31 (refs 9, 16). These data should also apply to mouse DHFR as the amino-acid sequences of the two enzymes show 77% homology^{17,18}. Such extensive sequence homology predicates close resemblance of the three-dimensional structures⁹.

It could be argued that methotrexate by itself might not be able to pass membranes and thereby might block translocation of any protein to which it is bound. This possibility is rendered highly unlikely by the finding that 1,3-diamino-7-benzylpyrrolo-[3,2-*f*]-quinazoline also blocks import of the purified fusion protein; this folate antagonist is uncharged and highly hydrophobic and has been shown to pass freely through cell membranes of yeast¹⁰ and *E. coli* (R.L. Then, personal communication).

The combined evidence strongly suggests that folate antagonists block the translocation of DHFR through mitochondrial membranes by preventing unfolding of the protein. This view is in accord with the crystallographic evidence on the interaction between antifolates and DHFR and the protection data described here. By binding tightly to DHFR, antifolates impede unfolding of the enzyme. In line with this interpretation is the recent observation of a translocation intermediate during the import of the precursors of the $\text{F}_1\text{-ATPase}$ β -subunit and cytochrome c_1 into *Neurospora* mitochondria¹⁹. This intermediate accumulated at low temperature and appeared to span both mitochondrial membranes, as it remained accessible to an externally added protease even though the presequence had already been cleaved off by the matrix-located processing protease. To allow such an orientation, at least the amino-terminal portion of the imported proteins must have been partially unfolded. Neither of these observations, however, give any indication as to how extensive the unfolding must be to allow movement of the protein across a membrane. The purified, translocation-competent fusion protein described here should be a valuable tool for investigating this question.

This study was supported by grants 3.394-0.83 and 3.660-0.84 from the Swiss NSF. We thank Dr J. Kraut (University of California, San Diego) for sending us the unpublished crystallographic coordinates of chicken DHFR; Drs R. L. Then and D.

Stüber (Hoffman-La Roche Co., Basel) for gifts of antifolates and mouse DHFR, respectively; Dr E. Mehler (Biocenter, Basel) for valuable advice; W. Oppliger for technical assistance; and M. Probst and L. Schwörer for typing the manuscript.

Received 1 April; accepted 22 May 1986.

1. Wickner, W. T. & Lodish, H. F. *Science* **230**, 400-407 (1985).
2. Blobel, G. & Dobberstein, B. *J. Cell Biol.* **67**, 835-851 (1975).
3. Randall, L. L. & Hardy, S. J. S. in *Modern Cell Biology* Vol. 3, 1-20 (Liss, New York, 1984).
4. Cullis, P. R. & De Kruijff, B. *Biochim. biophys. Acta* **559**, 399-420 (1979).
5. Hurt, E. C., Pesold-Hurt, B. & Schatz, G. *FEBS Lett.* **178**, 306-310 (1984).
6. Hurt, E. C., Pesold-Hurt, B., Suda, K., Oppliger, W. & Schatz, G. *EMBO J.* **4**, 2061-2068 (1985).
7. Chang, A. C. Y. *et al. Nature* **275**, 617-624 (1978).
8. Kaufman, B. T. & Pierce, J. V. *Biochem. biophys. Res. Commun.* **44**, 608-613 (1971).
9. Volz, K. W. *et al. J. biol. Chem.* **257**, 2528-2536 (1982).
10. Castaldo, R. A., Gump, D. W. & McCormack, J. J. *Antimicrob. Ag. Chemother.* **15**, 81-86 (1979).
11. Maarse, A. C. *et al. EMBO J.* **3**, 2831-2837 (1984).
12. Young, E. T. & Pilgrim, D. B. *Molec. cell. Biol.* **5**, 3024-3034 (1985).
13. Van Loon, A. P. G. M., Brändli, A. & Schatz, G. *Cell* **44**, 801-812 (1986).
14. Hurt, E. C., Pesold-Hurt, B. & Schatz, G. *EMBO J.* **3**, 3149-3156 (1984).

15. Böhni, P. C., Daum, G. & Schatz, G. *J. biol. Chem.* **258**, 4937-4943 (1983).
16. Matthews, D. A. *et al. J. biol. Chem.* **260**, 381-391 (1985).
17. Kumar, A. A., Blankenship, D. T., Kaufman, B. T. & Freisheim, J. H. *Biochemistry* **19**, 667-678 (1980).
18. Stone, D. & Phillips, A. W. *FEBS Lett.* **74**, 85-87 (1977).
19. Schleyer, M. & Neupert, W. *Cell* **43**, 339-350 (1985).
20. Dretzen, G., Bellard, M., Sassone-Corsi, P. & Chambon, P. *Analyt. Biochem.* **112**, 295-298 (1981).
21. Amann, E., Brosius, J. & Ptashne, M. *Gene* **25**, 167-178 (1983).
22. Bolivar, F. & Backman, K. *Meth. Enzym.* **68**, 245-268 (1979).
23. Stanley, K. K. *Nucleic Acids Res.* **11**, 4077-4092 (1983).
24. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
25. Neu, H. C. & Heppel, L. A. *J. biol. Chem.* **240**, 3685-3692 (1965).
26. Gupta, S. V. *et al. Biochemistry* **16**, 3073-3079 (1977).
27. Wray, W., Boulakas, T., Wray, V. P. & Hancock, R. *Analyt. Biochem.* **118**, 197-203 (1981).
28. Van Loon, A. P. G. M., Maarse, A. C., Riezman, H. & Grivell, L. A. *Gene* **26**, 261-272 (1983).
29. Gasser, S. M., Daum, G. & Schatz, G. *J. biol. Chem.* **257**, 13034-13041 (1982).

LETTERS TO NATURE

Infrared helioseismology: detection of the chromospheric mode

Drake Deming*, David A. Glenar†, Hans Ulrich Käuff*, Andrew A. Hill† & Fred Espenak*

* Planetary Systems Branch, NASA/Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

† Department of Physics, Colgate University, Hamilton, New York 13346, USA

Observations of velocity waves from the 5-min p-mode oscillations in the solar sub-photospheric cavity have permitted a number of important inferences about physical conditions and differential rotation in the solar interior¹. Models also predict the existence of an overlying acoustic cavity, defined by reflections between the temperature minimum and the sharp rise from chromospheric to coronal temperatures²⁻⁵. Observations of chromospheric emission features^{6,7}, the cores of strong absorption lines⁸⁻¹¹ and the far-infrared continuum¹² have established that chromospheric oscillatory power is shifted to higher frequencies than is the case for the sub-photospheric cavity. However, it remains to be established that this oscillatory power is sufficiently localized to indicate the presence of resonant modes at frequencies close to the predicted eigenfrequencies. We have obtained time-series observations of an infrared (11- μ m) solar OH absorption line profile, on two consecutive days, using a laser heterodyne spectrometer to view a 2-arc-s portion of the quiet Sun at disk centre. A power spectrum of the line-centre velocity shows the well-known photospheric p-mode oscillations very prominently, but also shows a second feature near 4.3 mHz. A power spectrum of the line intensity shows only the 4.3-mHz feature, which we identify as the fundamental ($n=1$) p-mode resonance of the solar chromosphere. The frequency of the mode is observed to be in substantial agreement with the eigenfrequency of current chromospheric models. A time series of two beam difference measurements shows that the mode is present only for horizontal wavelengths >19 Mm. The period of a chromospheric p-mode resonance is directly related to the sound travel time across the chromosphere, which depends on the chromospheric temperature and geometric height. Thus, detection of this resonance will provide an important new constraint on chromospheric models.

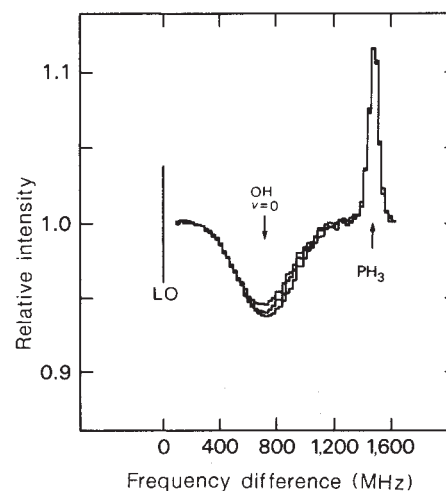


Fig. 1 Sample line profiles obtained on 16 June 1985 in three consecutive integrations, showing changes in the solar line profile and stability in the PH₃ gas-cell line. LO, local oscillator (¹³C¹⁶O₂ P(12), 903.74974 cm⁻¹); OH, R22(27.5)f rotational transition of OH, PH₃, phosphine absorption feature.

We measured the Doppler shift and line intensity of a 903.769-cm⁻¹ (11.065- μ m) OH absorption feature with a 2-arc-s field of view (full width at half-maximum, FWHM) centred on the solar disk, and not following the solar rotation. We used the Goddard laser heterodyne spectrometer at the McMath telescope of the National Solar Observatory, in June 1985. The design and operation of the instrument are discussed in ref. 13. The R22(27.5)f rotation transition of OH was measured using a peak-power stabilized laser local oscillator operating on the I-band, P(12) transition of ¹³C¹⁶O₂ at 903.7497 cm⁻¹. The observations were made in the optically null-balanced mode described in ref. 14. In this mode, the instrument uses a 27-Hz mirror-chopper to alternately place the Sun and a local high-temperature black body on the detector. Line profiles are measured in the optical difference spectrum. To monitor the laser stability, an absorption cell containing phosphine (PH₃) at 0.5 torr was placed in the black-body (BB) beam. A PH₃

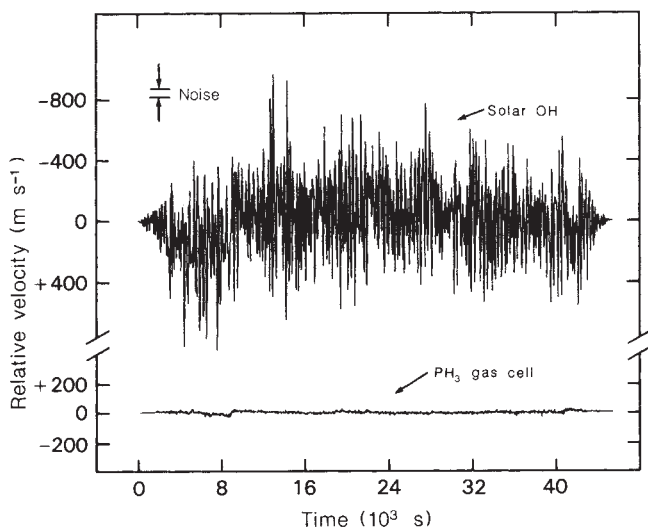


Fig. 2 Time series of line-centre velocity for the solar OH feature and PH_3 gas-cell line. The velocity scales are of opposite sense because the lines occur in opposite side bands. Both time series were apodized using a 10% cosine bell. The 1σ statistical noise level for a single solar measurement is 26 m s^{-1} and is indicated in the upper left. The corresponding 1σ level for the PH_3 line is 5 m s^{-1} .

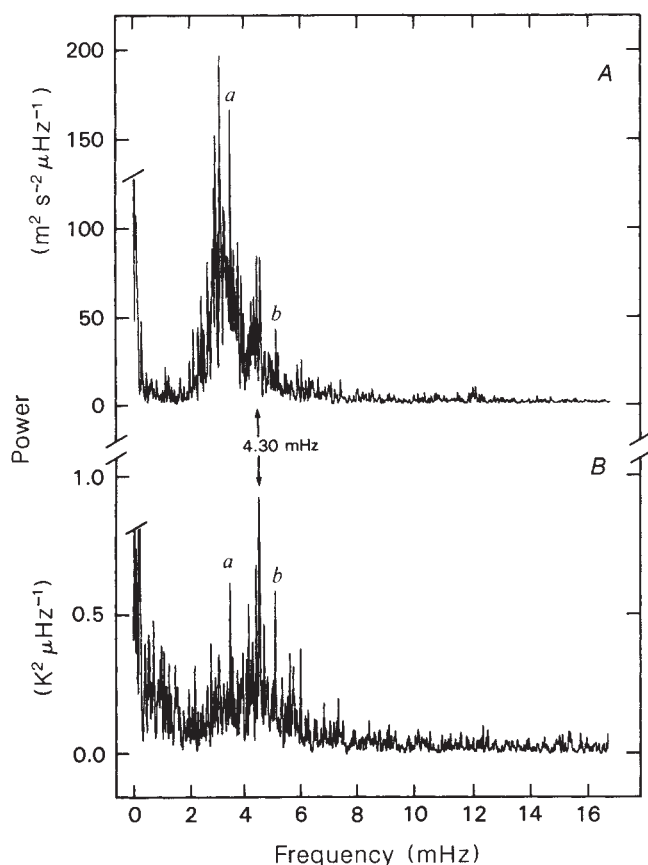


Fig. 3 Power spectra of line velocity (A) and line intensity (B), calculated from the concatenated data of 15 and 16 June, and smoothed to $20 \mu\text{Hz}$ resolution. The line velocity was measured in metres per second and the line depth was measured in kelvins. The peaks near zero frequency in both spectra are due to instrumental effects, and are not of solar origin. The arrows indicate the solar chromospheric mode at 4.30 mHz , and the features labelled 'a' and 'b' are discussed in the text.

absorption feature is thus recorded simultaneously with the solar OH feature in each 30-s integration. In Fig. 1, three consecutive integrations show the solar OH feature near the centre, and the PH_3 feature near $1,450 \text{ MHz}$. The PH_3 line appears inverted because each result is obtained as a difference spectrum (Sun-BB). A gaussian profile is fitted to each line to determine the line centre position, and line intensity (depth) relative to the continuum. It is evident from Fig. 1 that the solar line shows large fluctuations in both line centre frequency and line intensity. These fluctuations are not seen in the PH_3 gas-cell line, indicating that potential instabilities in the local oscillator frequency are negligible.

Our Sun-centre time series spanned a total of 2 days (15–16 June), with a total of 18.7 h of data. Figure 2 shows apodized velocity time series for 16 June. The results of the PH_3 profile fitting are displayed below the solar results to show the overall instrument stability. The solar time series shows r.m.s. (root-mean-square) velocity fluctuations of $\sim 300 \text{ m s}^{-1}$, greatly in excess of the 26 m s^{-1} (1σ) error associated with a single measurement. The only instrumental effect which is noticeable in Fig. 2 is represented by a slow drift in both time series, which is corrected near $8,600 \text{ s}$. This is caused by changes in the spectral baseline due to changes in the detector response. The least-squares algorithms which compute the line shape parameters cannot fully compensate for these baseline drifts, because the bandpass of the heterodyne spectrometer is so narrow. However this instrumental effect has no impact for the present purpose, because it propagates into the power spectra only at the lowest frequencies.

Velocity and line intensity power spectra from both days of observations were calculated using the method of Deeming¹⁵: the velocity spectrum is shown in Fig. 3A. The spectrum of the concatenated data has been smoothed to a resolution of $20 \mu\text{Hz}$ (FWHM) to suppress the side-lobes produced by the night-time data gap. We checked the extent of side-lobe contamination by calculating power spectra of continuous sine waves with the same data sampling. The velocity spectrum shows the dominant 3-mHz band from the subphotospheric cavity. Individual modes in this band cannot be distinguished because of the lack of spatial wavenumber resolution. The second most prominent feature is the relatively narrow band of power near 4.3 mHz , which we identify as a chromospheric resonance. The integrated power in the range $3.86\text{--}4.47 \text{ mHz}$ produces an r.m.s. velocity amplitude of 50 m s^{-1} . This is too large to be attributed to instrumental drifts or solar image motion at this frequency. The reality of the 4.3-mHz feature is confirmed in Fig. 3B, which shows the power spectrum of the line depth. In this figure the photospheric p-modes are essentially absent; only the 4.3-mHz feature is evident. The central peak in this spectrum occurs at $4.30 \pm 0.01 \text{ mHz}$, in very close agreement with the locus of frequencies predicted by Ando and Osaki³ for radial order $n = 1$. Two other peaks, labelled 'a' and 'b', are also prominent in both spectra at identical frequencies. They occur at 3.39 mHz (a) and 4.96 mHz (b). Although they do not correspond to overtones of higher radial order ($n > 1$), we speculate that they may represent resonances due to different components in an inhomogeneous chromosphere.

Several factors lend confidence to our identification of the $n = 1$ chromospheric resonance. (1) Our infrared observations sample the upper photosphere, where the chromospheric mode becomes increasingly prominent relative to the photospheric mode¹⁶. (2) The mode is observed at a frequency in close agreement with the predicted eigenfrequency. (3) The mode is clearly present in power spectra of the line intensity as well as the line velocity. It is distinguished from the photospheric p-modes by being relatively more prominent in the line intensity spectrum. This is expected, as the temperature perturbation associated with the chromospheric mode will be less affected by radiative damping¹⁷. (4) Compared with many previous chromospheric observations, our data are of relatively long

duration, making our result more representative of average conditions in the upper solar atmosphere. (5) The lack of solar activity at the time of our observations provided optimum conditions for the existence of a resonance.

In addition to the power spectra of Fig. 3, we have also examined power spectra from a 1-day sequence of dispersion-line shapes, obtained on 18 June 1985 by chopping our 2-arc-s field of view between two locations on the solar disk separated by 26 arc s (see ref. 13). These data allow us to derive the velocity and line intensity difference between these two locations, and power spectra of these quantities show no evidence of the chromospheric mode. This implies that the mode is being coherently subtracted in these data; therefore, its horizontal wavelength must exceed the 19-Mm chopping distance. A long horizontal wavelength for this mode is also consistent with the low-spatial-resolution data of Lindsey and Kaminski¹², provided that we identify the chromospheric mode as producing the 3–6-mHz oscillatory power which they observed in the intensity

of the far-infrared solar continuum. This situation was anticipated by Christensen-Dalsgaard and Frandsen⁵, who commented that "modes whose horizontal wavelength is much longer than the scale of the inhomogeneities feel only the mean structure of the atmosphere, ... therefore our results may apply to such modes".

We thank J. Goldstein for assistance in the observations, J. Faris for preparation of CO₂ mixtures, the staff of the National Solar Observatory for their support of our experiment, and our colleagues Drs T. Kostiuik and M. J. Mumma for their efforts in the development of the laser heterodyne technique at Goddard. D.A.G. was supported by NASA/Goddard Space Flight Center as a NASA/ASEE summer faculty fellow, and by NASA/Goddard Space Flight Center grant NAG 5-478. The present investigation was supported by RTOP 188-41-55 to Dr Mumma. A.A.H. received support from Colgate University's Division of Natural Sciences; D.D., D.A.G. and H.U.K. were guest investigators at the National Solar Observatory.

Received 10 February; accepted 23 April 1986.

1. Christensen-Dalsgaard, J., Gough, D. & Toomre, J. *Science* **229**, 923–931 (1985).
2. Bahng, J. & Schwarzschild, M. *Astrophys. J.* **137**, 901–913 (1963).
3. Ando, H. & Osaki, Y. *Publ. astr. Soc. Japan* **29**, 221–233 (1977).
4. Ulrich, R. K. & Rhodes, E. J. Jr *Astrophys. J.* **218**, 521–529 (1977).
5. Christensen-Dalsgaard, J. & Frandsen, S. *Sol. Phys.* **82**, 165–204 (1983).
6. Orrall, F. Q. *Astrophys. J.* **143**, 917–927 (1966).
7. Athay, R. G. & White, O. R. *Astrophys. J. Suppl. Ser.* **39**, 333–346 (1979).
8. Cram, L. E. *Astr. Astrophys.* **70**, 345–354 (1978).
9. Kneer, F., Newkirk, G. Jr & von Uexkull, M. *Astr. Astrophys.* **113**, 129–134 (1982).

10. Dame, F., Gouttebroze, P. & Malherbe, J. M. *Astr. Astrophys.* **130**, 331–340 (1984).
11. von Uexkull, M., Kneer, F., Mattig, W., Nesis, A. & Schmidt, W. *Astr. Astrophys.* **146**, 192–194 (1985).
12. Lindsey, C. & Kaminski, C. *Astrophys. J.* **282**, L103–L106 (1984).
13. Glenar, D., Deming, D., Espenak, F., Kostiuik, T. & Mumma, M. J. *Appl. Opt.* **25**, 58–62 (1986).
14. Deming, D., Hillman, J. J., Kostiuik, T., Mumma, M. J. & Zipoy, D. M. *Sol. Phys.* **94**, 57–74 (1984).
15. Deeming, T. J. *Astrophys. Space Sci.* **36**, 137–158 (1975).
16. Leibacher, J., Gouttebroze, P. & Stein, R. F. *Astrophys. J.* **258**, 393–403 (1982).
17. Noyes, R. W. & Leighton, R. B. *Astrophys. J.* **138**, 631–647 (1963).

Shape memory behaviour in partially stabilized zirconia ceramics

M. V. Swain

CSIRO Division of Materials Science, Locked Bag 33,
Normanby Road, Clayton, Victoria 3168, Australia

Shape memory behaviour has been observed in a variety of metallic alloys. It is usually associated with specific materials which on being deformed to various degrees of inelastic strain, return to their original shape on heating. The essential requirement is that they undergo reversible martensitic (diffusionless) phase transformations on heating and cooling. The past decade has seen considerable scientific and technological interest in ceramics containing zirconia (ZrO₂), a material that exhibits such martensitic transformations. This interest has arisen because of the high strengths and toughness that can be achieved in such materials by a stress-induced volume-expanding phase transformation about the crack tip¹. Observations are reported here of shape memory behaviour in a zirconia alloy partially stabilized with magnesia, at temperatures a few hundred degrees higher than reported for metallic alloys. The reversible deformation strains are relatively small, 0.5%, and are limited by the brittleness of the ceramic material. In the MgO-PSZ material the phase exhibiting the martensitic behaviour is the minor phase and occurs as precipitates in a stable matrix.

In the course of thermal cyclic testing of magnesia-partially-stabilized-zirconia (MgO-PSZ) material (J. E. Excel, M. Marmach and M. V. S., unpublished results) a behaviour very similar to the so-called shape memory effects of certain metallic alloys was observed. Shape memory behaviour was first observed about 40 years ago in various metallic alloys, and has since found technological applications². The metallic alloys based on compositions containing NiTi or CuAlNi have attracted much interest. The copper-containing alloy has created interest recently through an attempt to increase the temperature at which the deformation remains reversible³.

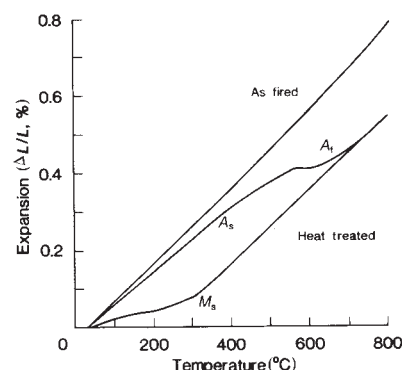


Fig. 1 Thermal expansion curves for the as-fired and heat-treated MgO-PSZ.

The essential requirement for shape memory alloys is that they undergo reversible martensitic phase transformations on heating and cooling, and this usually also means that the twin boundaries of the lower-symmetry (lower-temperature) phase must be sufficiently mobile or glissile to enable reorientation on the application of stress. The twin boundaries within the low-temperature phase must then consist of partially or fully coherent interfaces, to enable reversal on heating. Deformation of the alloy under load usually takes place below the martensitic start temperature, M_s , with non-linear strain being accommodated by the reorientation of the twin laths within the microstructure. It is also possible to achieve such non-linear strain above M_s but below a specific temperature, M_d , by a stress-induced transformation. On heating materials that have been deformed in either manner to above the A_f temperature, where the lower-temperature crystal structure completely reverts back to the high-temperature phase, the bodies return to their original pre-strained states. The reversal of the deformation begins at the temperature at which the reverse martensitic phase change starts, A_s , and is complete at A_f . Shape memory behaviour has usually been described for metals where the entire alloy is single phase, although some data exist for NiTi alloys containing minor volume fractions of intermetallic precipitates².

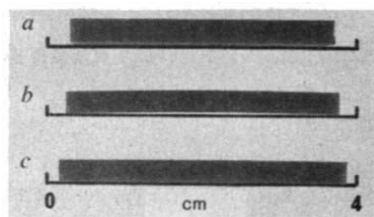


Fig. 2 Optical micrographs of the as-fired (*a*) and heat-treated (*b*) material after thermal cycling to 800 °C under a flexural stress of 200 MPa. *c*, Thermal cycling of a deformed bar to 800 °C in the absence of stress leads to the near complete reversal of the deflection.

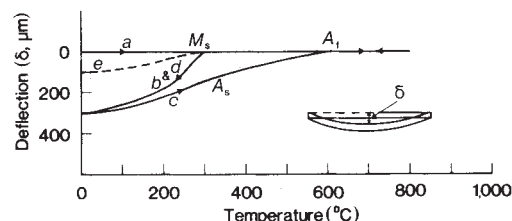


Fig. 3 A plot of the non-elastic deflection of the heat-treated material on heating and cooling under stress (corrected for thermal expansion of as-fired material). *a*, Heating under load (200 MPa); *b*, cooling under load (200 MPa); *c*, re-heating under load (200 MPa); *d*, cooling under load (200 MPa); *e*, cooling without load.

PSZ ceramics are a new class of strong and tough materials that may be fabricated to use a martensitic transformation to increase their toughness¹. The tetragonal phase of zirconia undergoes a martensitic transformation to monoclinic zirconia, accompanied by a 4–6% volume dilation and 9% shear strain. The temperature at which this transformation takes place is dependent on the solute content, precipitate size and matrix constraint, if any, imposed on the tetragonal material. To optimize the toughness of these materials, the content and size of the tetragonal grains or precipitates is adjusted so that M_s is just below the temperature of interest. The application of tensile stresses, for example about a crack tip, can trigger the volume-expanding transformation, leading to crack-tip shielding and a higher fracture toughness. With this mechanism a 10-fold increase in K_{Ic} is achievable in optimum alloy compositions⁴. In addition to the increased toughness, recent studies have shown that the application of stress may lead to ductile behaviour due to the tetragonal to monoclinic induced transformation⁵, and deformation strains of 0.25% have been reported^{5–7}.

The present observations were noted during thermal cycling tests on MgO-PSZ alloys in four-point flexure. These materials, containing 9.4 mol % MgO, consisted of coarse grains ($\sim 50 \mu\text{m}$) of a cubic zirconia matrix with ~ 35 vol. % tetragonal zirconia precipitates in the as-fired state. Some of these bars were heat-treated (at 1,100 °C for 9 h) to partially destabilize the tetragonal precipitates on cooling; that is, M_s was above room temperature. Example of the thermal expansion curves of the as-fired and heat-treated materials are shown in Fig. 1. The M_f temperature of the aged material was below room temperature and the monoclinic content of the polished surface as determined from the 111 X-ray diffraction peaks⁸ was 15%. Details of the heat treatment of MgO-PSZ materials are discussed elsewhere⁹. The thermal expansion hysteresis of PSZ alloys is much larger, particularly in the temperature axis, than most metallic alloys, and has been attributed to the constrained nature of the transforming precipitates¹⁰.

Thermal cycling tests consisted of loading rectangular bars, $3 \times 4 \times 40$ mm, in a four-point flexure, dead-weight loading jig, such that the tensile stress at the outer fibre was 200 MPa. No difference in behaviour was detected between specimens that had a ground or polished surface. The inner and outer spans of the flexure jig were 15 mm and 30 mm. The bars were slowly heated at 100°C h^{-1} to 800 °C and held for 2 h before cooling at 100°C h^{-1} . The as-fired material on such a thermal cycle behaved as anticipated for an elastic non-creeping body, whereas the heat-treated material was quite bent after such a cycle. A comparison of the two bars after such a thermal cycle under load is shown in Fig. 2*a, b*. On re-heating the bent bar to 800 °C at the same heating and cooling rates in the absence of stress, it returned almost completely to its original shape, as seen in Fig. 2*c*.

The permanent strain in the outer fibre of the bent bar after thermal cycling under load was calculated from the radius of

curvature, ρ , using the simple relationship

$$\epsilon = t/2\rho \quad (1)$$

where t is the beam thickness. This analysis assumed that the neutral axis remains in the middle of the deformed bar. Recent studies of stress-induced transformation during flexure have shown that the neutral axis is displaced⁶, and the above estimate for the strain is slightly conservative. From equation (1) the retained deformation strain at room temperature was found to be 0.42%, which is much larger than the deformation strains (typically 0.2%) reported in peak-toughness MgO-PSZ and CeO₂-PSZ alloys loaded in tension or flexure^{5–7}. The retained inelastic strain after re-heating to 800 °C and cooling to room temperature in the absence of external stress was 0.14%. The influence of applied stress on the extent of deformation strain and retained strain after re-heating in the absence of stress is being investigated and will be reported elsewhere.

To determine when the deformation was taking place, a linear variable displacement transducer (LVDT) was used to monitor the movement of the loading arm during heating and cooling. On heating and cooling the loading arms expanded and contracted substantially more than the specimen deformation. However, by subtracting the thermal expansion and contraction of the as-fired material from that of the heat-treated material it was possible to deduce the temperature at which deformation began. An additional experiment of re-heating the bent specimen under load was also performed to evaluate if, and at what temperature, the deformation reversed. The corrected deflection of the bar with temperature during heating and cooling on the first and second thermal cycles is shown in Fig. 3. Also shown are the M_s , A_s and A_f temperatures determined from the thermal expansion curve in Fig. 1.

The results in Fig. 3 show that the bar begins to deform on cooling at temperatures $\leq M_s$. On re-heating under load the bar begins to return to its original shape at temperatures below A_s and the process seems to be complete by A_f . On cooling, the deformation is identical to the initial cycle at temperatures below M_s . These observations are very similar to the observations in metallic shape memory alloys. The main differences are that the temperatures at which the deformation and reversal occur are hundreds of degrees higher in the MgO-PSZ and the extent of deformation strain is much less. There appears to be a small remnant of irreversible deformation associated with reheating above A_f in the absence of stress. This behaviour is also observed in metallic alloys, which exhibit a large hysteresis between M_s and A_f (ref. 2).

Crystallographic analysis of the orientation relationships of transformed monoclinic zirconia precipitate in a cubic matrix is an area of active research. Recent detailed transmission electron microscopy (TEM) of the twinned monoclinic precipitates in MgO-PSZ has shown that 12 different crystallographic variants exist¹¹. It has also been observed that the twin orientations in the monoclinic structure could differ from precipitate to precipitate and were dictated by the local stress state.

Microcracks, observed at the interface between matrix and precipitate, accommodated lattice strains which arose, particularly at specific twin domain/interphase boundaries. More recently, Hannink *et al.*¹² used a technique of *in situ* TEM straining of almost identical MgO-PSZ alloys using electron-beam heating and subsequently annealing transformed precipitates in the same manner. They also observed that the orientation of the monoclinic twin variants of individual precipitates is dependent on the superimposed local stress state. On annealing, the precipitates revert back to their original tetragonal shape, regardless of the pre-existing twin variants. In these TEM experiments, there was no attempt to monitor the temperature of the foil and therefore determine the M_s and A_s of individual precipitate. However, some residual strain did remain after annealing, particularly at sites where microcracking was observed when the precipitate was in the twinned monoclinic form. The presence of such residual strain sites may well contribute to the destabilization of some of the associated metastable tetragonal precipitates in a preferred manner on cooling. This would lead to some residual strain after re-annealing in the absence of a biasing or superimposed stress on the body, and may account for the slight residual deflection of the beam.

The present experiments strongly support the contention that shape memory behaviour, due to a martensitic reaction, occurs in PSZ ceramics. The application of a biasing stress as the body cooled through the M_s temperature leads to a deflection of the specimen which could be almost completely removed on annealing above the A_f temperature. Although the present experiments have been carried out with MgO-PSZ alloys, it is anticipated that similar behaviour would be observed in other zirconia-toughened ceramics with the appropriate phase assemblage. [Since submitting this manuscript, shape memory behaviour has also been observed in a fine-grained yttria (Y_2O_3)-PSZ ceramic (M. Matsui and T. Soma, personal communication.) At this stage it is uncertain what contribution the volume dilation associated with the transformation makes to the deformation strain developed in cooling through M_s under load. In addition, the ability of the stiff non-transformable cubic matrix phase to absorb the large deformations during cooling and heating through M_s and A_s under load is not understood. The limited strains associated with the inelastic deflection and subsequent recovery on annealing with the investigated MgO-PSZ alloy are so small that it is too early to speculate about technological applications of such shape memory behaviour. However, the much higher temperatures at which this behaviour is being observed in PSZ materials, and the ability to vary the M_s and A_f temperatures to $>1,200^\circ\text{C}$ by the addition of hafnium oxide, may provide technological opportunities unavailable for metallic shape memory alloys. The upper temperature application limit for zirconia ceramics, like metallic alloys, will be controlled by the temperature at which the internal elastic strains responsible for the shape memory effect are relaxed by thermally activated processes.

The provision of specimens by J. Excel and M. Marmach is appreciated, as are discussions with and access to unpublished work of R. H. J. Hannink, L. R. F. Rose, N. Kennon, D. Dunne and M. J. Bannister. I thank H. G. Scott and Professor A. P. Miodownik for constructive comments.

Received 28 November 1985; accepted 2 May 1986.

- Garvie, R. C., Pascoe, R. T. & Hannink, R. H. J. *Nature* **258**, 703-705 (1975).
- Metals Forum* **4**, 125-190 (1981).
- Duerig, T. W., Albrecht, J. & Gessinger, G. H. J. *Metals* **32**, 14-20 (1982).
- Swain, M. V. & Rose, L. R. F. *Advances in Fracture Research*, ICF6 Vol. 1 (eds Valluri, S. R. *et al.*) 473-494 (Pergamon, Oxford).
- Swain, M. V. *Acta metall.* **33**, 2035 (1985).
- Marshall, D. B. J. *Am. ceram. Soc.* **69**, 173-180 (1986).
- Tsukuma, K. & Shimada, M. *J. mater. Sci.* **20**, 1178-1184 (1985).
- Garvie, R. C. & Nicholson, P. *Bull. Am. ceram. Soc.* **55**, 303-305 (1972).
- Hannink, R. H. J. & Swain, M. V. *J. Aust. ceram. Soc.* **18**, 53-62 (1982).
- Hannink, R. H. J., Johnson, K. A., Pascoe, R. T. & Garvie, R. C. *Adv. Ceram.* **3**, 116-136 (1981).
- Muddle, B. & Hannink, R. H. J. *J. Am. ceram. Soc.* (in the press).
- Hannink, R. H. J., Porter, R. J. & Marshall, D. B., *J. Am. ceram. Soc.* (in the press).

Unusual conditions in the tropical Atlantic Ocean in 1984

S. G. H. Philander

Geophysical Fluid Dynamics Laboratory/NOAA,
Princeton University, Princeton, New Jersey 08542, USA

During the first half of 1984, oceanic and atmospheric conditions in the tropical Atlantic were, in many respects, similar to conditions in the Pacific Ocean during El Niño: the upper ocean was unusually warm in the eastern part of the basin, rainfall was heavy over the normally arid regions to the south of the Equator, and coastal upwelling was inhibited in regions (southwestern Africa) where this is a seasonal phenomenon. The change in sea-surface temperatures, which resulted when unusual eastward currents to the south of the Equator transported warm surface waters towards Africa, contributed to the change in atmospheric conditions. The altered atmospheric conditions, in turn, contributed to the change in the oceanic circulation and the change in the sea-surface temperature.

Figure 1 contrasts the sea-surface temperatures (SSTs) in the tropical Atlantic in June 1983 and June 1984. The higher SSTs in 1984 were associated with exceptionally heavy rainfall over northeastern Brazil and the coastal zone of southwestern Africa but were also associated with a persistence of the drought in the Sahel, the region to the south of the Sahara desert. This happened because the Intertropical Convergence Zone¹ (ITCZ), the narrow east-west band of rising air, cloudiness, and heavy rainfall onto which the south-east and north-east trades converge, was displaced unusually far southward during 1984. Such a displacement of the ITCZ is also a feature of El Niño in the Pacific Ocean. However, the other important feature of El Niño, eastward movement of the atmospheric convective zone that is normally over the western equatorial Pacific, had no counterpart in the Atlantic during 1984. In the Pacific the eastward movement of the region of heavy rainfall and low sea-level pressure, towards the arid eastern equatorial Pacific where sea-level pressure is normally high, results in negatively correlated variations in the eastern and western sides of the ocean basin. During the warm event in the Atlantic, by contrast, the increase in rainfall and decrease in sea-level pressure was fairly uniform in the east-west direction.

The dominant change in the Atlantic was not a zonal one—the low-pressure convective zone over the Amazon Basin was not dislodged from the continent—but was a meridional one associated with the equatorward movement of the ITCZ. The ITCZ is the principal region of rainfall over the tropical Atlantic Ocean and the adjacent land. Seasonally it moves between the Equator and 15°N , approximately, so that northeastern Brazil has a rainy season in March and April when the ITCZ is furthest south, whereas the Sahel (10°N to 15°N in west Africa) has a rainy season centred on August when the ITCZ is furthest north. Because of this dependence on the ITCZ for rainfall, latitudinal precipitation gradients are enormous: annual mean rainfall along the west African coast is 1,939 mm at Bissau (12°N), 542 mm at Dakar (15°N) and 123 mm at Nouakchott (18°N). These statistics make it clear that slight variations in the position of the ITCZ, which has a very small north-south extent, have a major effect on rainfall variations in west Africa. The unusual southerly position of the ITCZ in 1984 contributed significantly to the persistence into 1984 of the drought in the Sahel, and to the heavy rainfall and floods in northeastern Brazil and the coastal zone of equatorial and southwestern Africa, including Angola.

Atmospheric convective zones over the ocean are usually over the warmest surface waters. The ITCZ, for example, is in the neighbourhood of the Equator when SSTs in that region are at

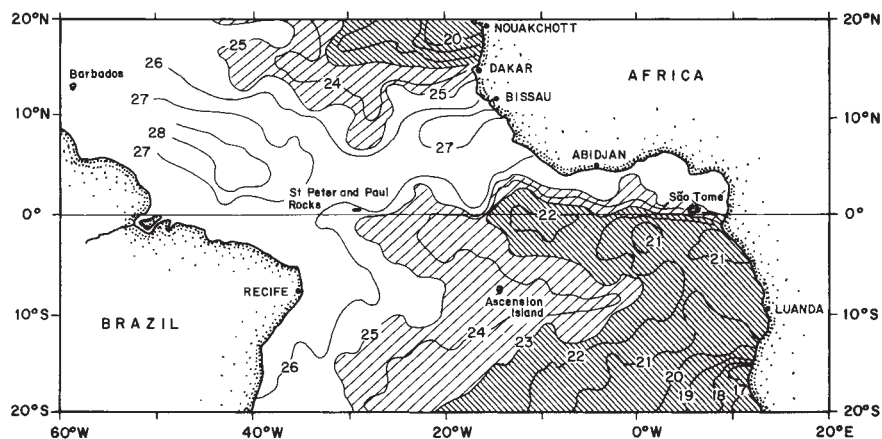
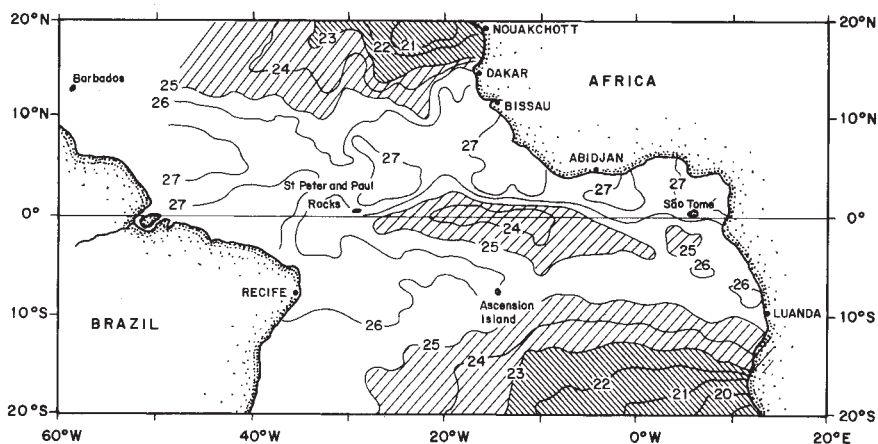


Fig. 1 Sea-surface temperatures (°C) in the tropical Atlantic Ocean in June 1983 (top) and June 1984 (bottom).



a maximum (March and April) and it moves northward when SSTs at, and to the south of the Equator in the central and eastern Atlantic start to fall. Interannually, southward displacement of the ITCZ is correlated with, and in models is effected by, an increase in SST to the south of the Equator and a simultaneous decrease in SST farther north^{2,3}. SST variations in the tropical Atlantic, therefore, influence rainfall variations over Africa and Brazil. There are, however, other factors that influence the position of the ITCZ and the rainfall. Palmer³ finds that, in a general circulation model of the atmosphere, SST variations in the tropical Pacific Ocean have a significant effect on the position of the ITCZ over the Atlantic Ocean. This result suggests that the atmospheric convergence onto the region of unusually high SST in the eastern tropical Pacific in 1983 was the cause of the intense trade winds over the Atlantic in 1983. Palmer's calculations indicate that the high SST in the Atlantic during 1984 should have weakened the trade winds over the Pacific. Such a weakening was not observed—the trade winds over the Pacific were stronger than usual in 1984—possibly because SST in the tropical Pacific during 1984 were much lower than normal. There are factors, other than SST, that influence the atmospheric circulation in the tropics. For example, it has been suggested that an expansion of the Northern Hemisphere circumpolar vortex and an equatorward displacement of the subtropical high-pressure belt affect the tropical circulation and the rainfall of the Sahel⁴⁻⁷. The degree to which SST variations in the tropical Atlantic affect rainfall in the Sahel appears to be modest⁸ but there are indications that these SST variations have a more pronounced effect on rainfall over northeastern Brazil and southwestern Africa.

SST variations over large parts of the ocean are determined primarily by changes in the local flux of heat across the ocean surface. In the tropical oceans, however, SST variations are

strongly influenced by changes in the large-scale oceanic circulation, in response to changes in the surface winds. On seasonal and interannual timescales the strength of the wind determines the intensity of the oceanic circulation and the magnitude of the associated thermal gradients in the ocean. These gradients are reflected in SST patterns. For example, SST is uniformly high in March and April when the winds are relaxed and drive a weak circulation. The seasonal strengthening of the winds in May intensifies the oceanic circulation and increases the horizontal thermal gradients both in the surface and subsurface. The strong westward surface flow to the south of 3°N now transports warm surface waters from the Gulf of Guinea in the east to the western side of the ocean basin so that SST is much lower in the southeastern equatorial Atlantic than in the west. The unusually weak surface winds in 1984 maintained relatively small thermal gradients in the tropical Atlantic Ocean, primarily because isotherms in the Gulf of Guinea were exceptionally deep during the early months of 1984 (refs 9-11). An unusual eastward current between the Equator and 5°S carried warm water into the Gulf of Guinea. Some of this water flowed southward along the African coast and suppressed coastal upwelling as far south as Angola and Namibia¹². Simulations of oceanic conditions in the tropical Atlantic during 1984 with a realistic general circulation model are in progress and should provide a coherent picture of how the circulation changed during the warm event.

From an oceanographic point of view, the unusual conditions in the tropical Atlantic Ocean during 1984 were caused by changes in the surface winds. From a meteorological point of view, the unusual winds were in part attributable to changes in the SST. This circular argument suggests that interactions between the ocean and atmosphere were of central importance. These interactions are crucial both seasonally and interannually.

(The large seasonal variations in SST near the Equator both influence the movements of the ITCZ and are influenced by the movements of the ITCZ which determine changes in the surface winds.) Interannual variations in the Atlantic Ocean can be viewed as perturbations to the seasonal cycle. In 1984, the seasonal migrations of the ITCZ took it further south than normal and it remained in a southerly position longer than normal. An unusual southward displacement of the ITCZ is also a feature of El Niño events in the Pacific Ocean. An additional feature of El Niño, but apparently not a feature of interannual variability in the Atlantic, is an eastward displacement of atmospheric convective zones at the western extreme of the ocean basin. Most studies¹³⁻¹⁵ of interactions between the ocean and atmosphere in the tropics concern this zonal movement of convective zones. Studies of the air-sea interactions and the other factors that control the latitudinal movements of the ITCZ will shed light not only on phenomena such as droughts in western Africa and northeastern Brazil, but also on El Niño.

Finally, how often do warm events such as that of 1984 occur? The previous warm event with a comparable amplitude to the south of the Equator occurred in 1963 when the arid regions of southwestern Africa experienced severe floods, and a suppression of the coastal upwelling. However, the few available long time-series records show that there is considerable variability

from year to year⁸. This variability is associated with the timing of the major feature of the seasonal cycle, the northward movement of the ITCZ. This movement starts abruptly so that the associated intensification of the south-east trades is sudden. In some (cold, dry) years, this happens as early as February, in some (warm, wet) years as late as June. Over the next decade, special oceanographic and meteorological measurements will be made in the tropical Atlantic to document this variability. This will be part of the international TOGA program, to study interannual variability of the tropical oceans and global atmosphere, over a 10-year period that started in 1985.

Received 17 January; accepted 18 April 1986.

1. Horel, J. D., Kousky, V. E. & Kagano, M. T. *Nature* **322**, 248-251 (1986).
2. Moura, A. D. & Shukla, J. *J. Atmos. Sci.* **38**, 2653-2675 (1981).
3. Palmer, T. N. *Nature* **322**, 251-253 (1986).
4. Winstanley, D. *Nature* **243**, 464-465 (1973).
5. Winstanley, D. *Nature* **245**, 190-194 (1973).
6. Angell, J. K. & Korshover, J. *Mon. Weath. Rev.* **102**, 669-678 (1974).
7. Lamb, P. J. *Tellus* **30**, 240-251 (1978).
8. Lamb, P. J., Pepler, R. A. & Hastenrath, S. *Nature* **322**, 238-243 (1986).
9. Hisard, P., Henin, C., Houghton, R., Piton, B., & Rual, P. *Nature* **322**, 243-245 (1986).
10. Katz, E. J., Hisard, P., Verstraete, J.-M. & Garzoli, S. L. *Nature* **322**, 245-247 (1986).
11. Weisberg, R. & Colin, C. *Nature* **322**, 240-243 (1986).
12. Shannon, L. V., Boyd, A. J., Brundrit, G. B. & Taunton-Clark, J. *J. mar. Res.* (in the press).
13. Philander, S. G. H., Yamagata, T. & Pacanowski, R. C. *J. Atmos. Sci.* **41**, 604-613 (1984).
14. Cane, M. A. & Zebiak, S. E. *Science* **228**, 1085 (1985).
15. Anderson, D. L. T. & McCreary, J. P. *J. Atmos. Sci.* **42**, 615-629 (1985).

Interannual variability in the tropical Atlantic

Peter J. Lamb*, Randy A. Pepler* & Stefan Hastenrath†

* Climate and Meteorology Section, Illinois State Water Survey, Champaign, Illinois 61820, USA

† Department of Meteorology, University of Wisconsin, Madison, Wisconsin 53706, USA

The anomalous meteorological and oceanic conditions in the tropical Atlantic basin during 1983-84 can be seen as an extreme example of a common historical pattern. Here we present time series of rainfall in subsaharan West Africa (SWA) and north-east Brazil (NEB), which show that departures in 1983-84 were severe, but not unprecedented. The associated sea surface temperature (SST) anomaly patterns included one that had previously been shown to accompany earlier years of extreme rainfall in SWA and NEB, and others that contained key elements of that pattern. The pattern concerned was recently shown to be a major mode of tropical Atlantic SST variation.

Figure 1 documents the interannual variation of rainfall in two contrasting parts of the tropical Atlantic basin, SWA (11-18° N) and NEB (3-8° S). Their rainy seasons, which occur at different times of the year (July-September core in SWA and March-April core in NEB), are associated with the two extremes of the annual cycle of the latitudinal position of the Intertropical Convergence Zone (ITCZ)¹. Each time series shown contains the yearly averages of normalized departures for a fixed set of stations (20 in SWA, 30 in NEB). The value for the year j is given by

$$X_j = \frac{1}{N_j} \sum_{i=1}^{N_j} \frac{r_{ij} - \bar{r}_i}{\sigma_i}$$

where r_{ij} is that year's rainfall total at station i , $\bar{r}_i$ and σ_i are respectively the mean and standard deviation of station i 's yearly totals for the base period used (1941-82 for SWA, 1913-56 for NEB), and N_j is the number of stations with complete records in year j . The statistical advantages of this rainfall index formulation were originally summarized by Kraus², and its viability has

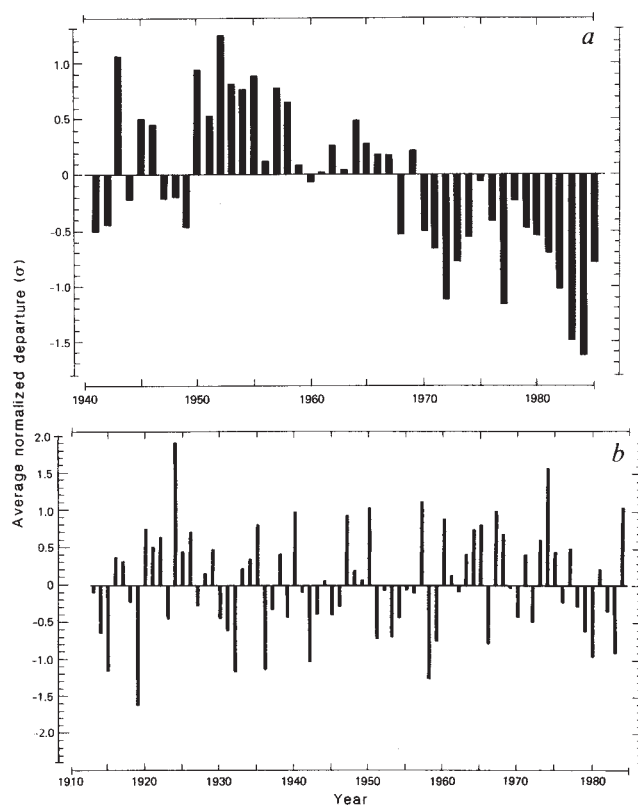


Fig. 1 Time series of yearly averages of normalized rainfall departures (in units of σ) for a, subsaharan West Africa (April-October) for 1941-85, and b, north-east Brazil (March-April) for 1913-84.

been confirmed by Katz and Glantz³. Further details on its use in this context appear in refs 4-6, where earlier versions of the time series were previously published for 1941-82 (SWA) and 1912-73 (NEB).

It is clear from Fig. 1 that 1983-84 was a period of anomalous rainfall in both SWA and NEB. Those two years were SWA's driest during 1941-84, and constituted the most recent extension

Fig. 2 Sea surface temperature (SST) anomaly maps ($^{\circ}\text{C}$) for January 1983 (*a*), August 1983 (*b*), March 1984 (*c*) and August 1984 (*d*), from ref. 13. The March 1983 map would have been a better indicator of conditions during the NEB rainy season than that for January 1983, but was not available.

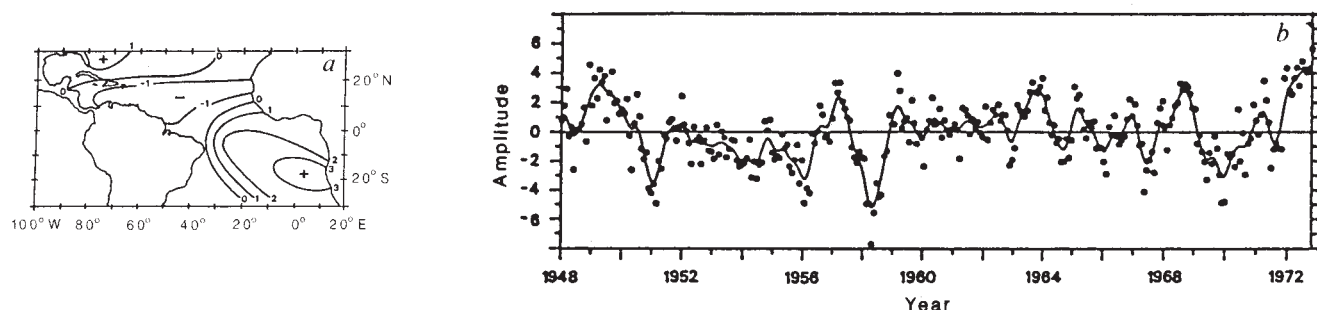
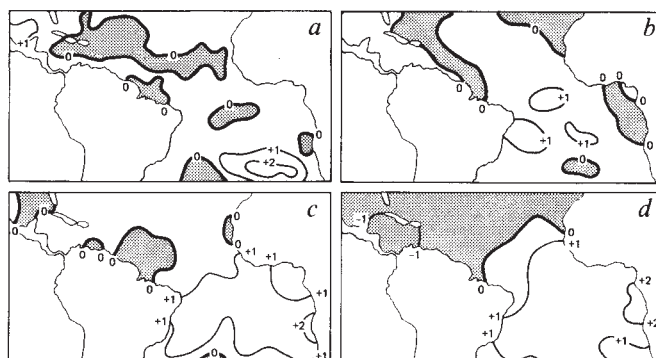


Fig. 3 Principal component 2 extracted from monthly normalized SST departures for 1948–72 (from ref. 12). *a*, Loadings multiplied by 10; *b*, amplitudes, with curve showing the result of applying a 13-point low-pass filter.

of a drought that has persisted, with very little respite, since 1968 (Fig. 1*a*). The widespread severity of the 1983–84 SWA drought conditions is fully documented in refs 7 and 8. Rainfall in that zone was also extremely deficient in 1972, 1977 and 1982, and moderately deficient during several other years since 1968, including 1985. Three moderately severe SWA drought years also occurred in the 1940s. The period spanning 1950–58, in contrast, was one of abundant rainfall in SWA, with strong positive departures in most of these years; weaker positive departures characterized 1959–67. SWA has thus experienced a pronounced decline of rainfall since the early 1950s.

In contrast to SWA, the anomalous 1983–84 rainfall in NEB consisted of both an extremely dry (1983) and an extremely wet (1984) year (Fig. 1*b*); each was among the 6–8 driest/wettest since 1913, and both ranked second in that regard for 1959–84. The stronger anomalies for this region are scattered throughout the record, which exhibits pronounced year-to-year variation. NEB rainfall shows little evidence of long-term trend, unlike that for SWA (Fig. 1). Hastenrath and Kaczmarczyk⁹ found that the variance of NEB rainfall includes a stronger high-frequency component (for example, periods of 2.5 and 5 yr) than does that of the Senegal river discharge, which reflects SWA rainfall. Much of the Senegal variance is centred on 30 yr, as is implied by the SWA time series in Fig. 1*a*, while NEB has a substantial portion of its interannual variance concentrated around 14 years. Preliminary reports (not shown) suggest that NEB rainfall was as abundant in 1985 as in 1984.

Figure 1 of ref. 10 documents the large-scale tropical Atlantic SST fields of June 1983 (a month lying between very deficient NEB and SWA rainy seasons; see Fig. 1 and discussion above) and June 1984 (lying between most abundant NEB and very deficient SWA rainy seasons). When converted to anomaly form, the 1984 pattern (see Fig. 2*c, d*) in particular is extremely similar to the one which has previously been shown to accompany earlier SWA drought and NEB flood years. That pattern, which was first identified by Lamb¹¹ (for SWA drought) and Hastenrath and Heller⁶ (for NEB flood), contains warmer than average surface water to the south of 10° N and east of 30° W, a feature that extends to at least 20° S, and colder than average surface

water to its immediate north and west. This pattern has been confirmed by the recent work of Lough¹², who showed it is a major mode of variation of tropical Atlantic SST (Fig. 3). The results of Lamb¹¹ and Lough¹² (including Fig. 3) collectively suggest that the above pattern has characterized many deficient SWA rainy seasons (for example, those in 1913, 1921, 1949, 1968, 1972 and, to a lesser extent, 1971), and did not occur during two periods (1927–34, 1950–56) when SWA was drought-free (see Fig. 1). Furthermore, the magnitude of the pattern's correlation with SWA rainfall, as obtained using Lough's¹² principal component amplitudes shown in Fig. 2, is in the range 0.51–0.66 (26–44% of rainfall variance explained), depending on the timescale of the analysis. Hastenrath and Heller⁶ found that the same anomaly pattern characterized an SST composite for 10 extremely wet NEB years, and that the opposite pattern emerged from a treatment of 10 extremely dry NEB years.

Although the above SST anomaly pattern was not fully present during the very deficient 1983 SWA rainy season (Fig. 2*b*), the strong positive departures centred on the Equator probably displaced the zone of maximum SST southwards of its long-term average position. Lamb¹¹ found the latter displacement to be an important consequence of the above SST anomaly pattern, and one that coincided with SWA drought. Note also that the positive SST departures in August 1983 were stronger between 5° N and 20° S than between 5° and 20° N. The SST anomaly field just before the deficient 1983 NEB rainy season (Fig. 2*a*) was also different in part from the NEB drought SST pattern described above. However, Fig. 2*a* does display the relatively large pool of abnormally cold water to the south of the Equator and the predominance of positive SST departures between 0 and 10° N. The 0–10° S SST anomaly pattern changed considerably from January to August of 1983 (Fig. 2*a, b*), in contrast to 1984 (Fig. 2*c, d*).

This research was partially supported by NOAA grants NA86AA-D-MC007 and NA86AA-D-AC065 and NSF grant ATM 84-13575. P.J.L. is also affiliated with the Department of Atmospheric Sciences, University of Illinois. We thank Phyllis Stone, Rebecca Runge and Linda Riggan for technical assistance.

Received 10 April; accepted 9 June 1986.

1. Hastenrath, S. *Mon. Weath. Rev.* **112**, 1097-1107 (1984).
2. Kraus, E. B. *Mon. Weath. Rev.* **105**, 1009-1018 (1977).
3. Katz, R. W. & Glantz, M. H. *Mon. Weath. Rev.* **114**, 764-771 (1986).
4. Lamb, P. J. *Nature* **299**, 46-48 (1982).
5. Lamb, P. J. *J. Clim.* **3**, 419-422 (1983).
6. Hastenrath, S. & Heller, L. *Q. Jl R. met. Soc.* **103**, 77-92 (1977).
7. Lamb, P. J. *Z. Gletscher. Glazialgeol.* **21**, 131-139 (1985).
8. Nicholson, S. E. *J. Clim. appl. Met.* **24**, 1388-1391 (1985).
9. Hastenrath, S. & Kaczmarczyk, E. B. *Tellus* **33**, 453-462 (1981).
10. Philander, S. G. H. *Nature* **322**, 236-238 (1986).
11. Lamb, P. J. *Tellus* **30**, 240-251 (1978); *Mon. Weath. Rev.* **106**, 482-491 (1978); *Clim. Monit.* **11**, 46-49 (1982).
12. Lough, J. M. *Mon. Weath. Rev.* **114**, 561-570 (1986).
13. *Climate Diagnostics Bulletin* (Climate Analysis Center, Washington, February 1983, August 1983, March 1984, August 1984).

Equatorial Atlantic Ocean temperature and current variations during 1983 and 1984

Robert H. Weisberg* & Christian Colin†

* Department of Marine, Earth, and Atmospheric Sciences,
North Carolina State University, Raleigh,
North Carolina 27695-8208, USA

† ORSTOM, 24 Rue Bayard, 75008 Paris, France

The equatorial regions of the Earth's oceans are climatically sensitive ones because of the zonal sea-surface temperature contrasts observed there¹. Equatorial sea-surface temperature normally varies on an annual cycle with the prevailing trade winds but deviations from this cycle may have significant global implications as occurred for example during the 1982-83 Pacific Ocean El Niño/Southern Oscillation event². Understanding the annual and interannual variability of the tropical oceans has therefore been the goal of several measurement programmes. In the Atlantic Ocean, the Seasonal Response of the Equatorial Atlantic (SEQUAL) Experiment and the Programme Français Ocean et Climat dans l'Atlantique Equatorial (FOCAL) have provided a basin-wide and synoptic data set over two annual cycles. We present here results from surface moored current meters which were one of several fixed and shipborne measurement systems employed by SEQUAL and FOCAL. We will describe the evolution of the upper ocean thermal and zonal velocity component variations in relation to forcing by the trade winds, show differences observed along the Equator at 28° W and 4° W, and compare the ocean responses at these locations during 1983 and 1984. The synoptic data realizations of these years differed from climatology and these differences are related to the rapidly varying nature and intensity of the wind stress in a given year. Changes in wind stress from year to year result in interannual variability as a modulated annual cycle and 1984, a year of weak winds relative to 1983, offers a case in point. The zonal sea-surface temperature gradient vanished along the Equator in 1984 during the season when it normally would have been a maximum.

The data were collected using vector averaging current meters suspended beneath taut moored surface buoys at depths of 10, 50, 75, 100, 150, and 200 m on the Equator at 28° W and at depths of 10, 35, 60, 85, and 110 m, on the Equator at 4° W. Sampling intervals for velocity and temperature were 15 min and these data have been low pass filtered to exclude fluctuations occurring at timescales shorter than 10 days to focus on the annual cycle.

Essentially, it is the zonal integral of easterly wind stress near the Equator that causes the annual cycle in the upper equatorial ocean velocity and temperature fields. Currents are generated which redistribute mass and together with equatorial long waves emanating from the boundaries these provide the mechanisms by which the ocean tends to adjust to the large-scale wind stress variations^{3,4}. Time series of surface wind velocity were collected during SEQUAL and FOCAL at St Peter and Paul Rocks (1° N,

29° W) and from buoys moored along the Equator at 24° W, 15° W and 4° W. The former are the most complete and they are representative of the temporal variations over the western two-thirds of the equatorial Atlantic⁵⁻⁷. Katz *et al.*⁸ show the details of the wind stress variability during 1983-84 and the following large-scale features are noted for qualitative discussion. During 1983 the easterly component of surface wind stress intensified rapidly beginning in mid-April and remained strongly developed through mid-December. It then relaxed rapidly and remained weak through mid-May before intensifying rapidly again. Immediately before the seasonal intensifications during both years were short lived (around one week) relaxation events where the already weak easterlies got even weaker or reversed to become westerlies.

The effects of these variations in easterly wind stress on the upper ocean thermal and eastward velocity component fields are apparent in the SEQUAL and FOCAL data. Figure 1 shows the temperature variability over the upper 200 m on the Equator at 28° W along with the temperature at 10 m which is approximately equal to sea-surface temperature (SST) and the vertically-integrated temperature which is proportional to heat content. Beginning in mid-April 1983, and coincident with the onset of increasing easterlies, a rapid shoaling of the isotherms and an attendant reduction in heat content is observed. This proceeds for around 1 month after which the isotherms deepen and the heat content increases to a relatively steady state lasting from August to December. Beginning in mid-December 1983 and coincident with the relaxation in easterly wind stress is a further deepening of the isotherms and an increase in heat content (peaking around mid-January) followed by shoaling through April. The thermocline has now completed one cycle in response to the winds and a repetition of this begins again during mid-May 1984 when the easterly wind stress intensifies again. The responses of the thermocline on the Equator at 28° W to the annual large-scale variations in easterly wind stress may, therefore, be summarized as sequences of shoaling and deepening (or conversely) leading to adjusted states. Accompanying these transitions are changes in the vertically-integrated temperature or heat content.

SST behaves differently to the thermocline depth. Note that while SST decreases to a minimum as the thermocline shoals beginning in mid-April 1983, it then remains low even after the thermocline has deepened to its adjusted state. In so doing, minimum SST on the Equator at 28° W coincides with a deepened thermocline and increased local heat content. The opposite occurs following the mid-December easterly wind stress relaxation resulting in maximum SST on the Equator at 28° W after the thermocline has shoaled and the local heat content has decreased.

Corresponding plots for the eastward velocity component, its vertical integral, and surface value are shown in Fig. 2. The flow is primarily eastward over the entire 200 m domain with only intermittent westward flow at the surface associated with the south equatorial current (SEC). The core of the equatorial undercurrent (EUC) is observed between 50 and 100 m and its position moves downward with the thermocline as the latter adjusts to increased easterly wind stress around June and July of both 1983 and 1984. While the vertical position of the EUC core migrates annually with the thermocline a clear annual variation in EUC speed or transport is not readily apparent.

Similar plots for the temperature and eastward velocity component observed on the Equator at 4° W are shown in Figs 3 and 4 respectively. The thermocline at 4° W also undergoes a sequence of shoaling followed by deepening beginning in mid-April 1983 and again in mid-May 1984, and a sequence of deepening followed by shoaling beginning in mid-December 1983. Several important differences between the responses at both longitudes are evident, however. First, the shoaling/deepening and deepening/shoaling sequences at 4° W are of longer duration and larger magnitude than those at 28° W.

Fig. 1 Thermal structure observed over the upper 200 m on the Equator at 28° W from February 1983 to October 1984. *a*, The temperature at 10-m depth; *b*, isotherm depths as a function of time; *c*, vertically integrated temperature proportional to heat content. The data have all been low pass filtered to remove fluctuations with timescales <10 days. Temperatures >27 °C are stippled to highlight the warmest surface temperatures relative to the thermocline depth.

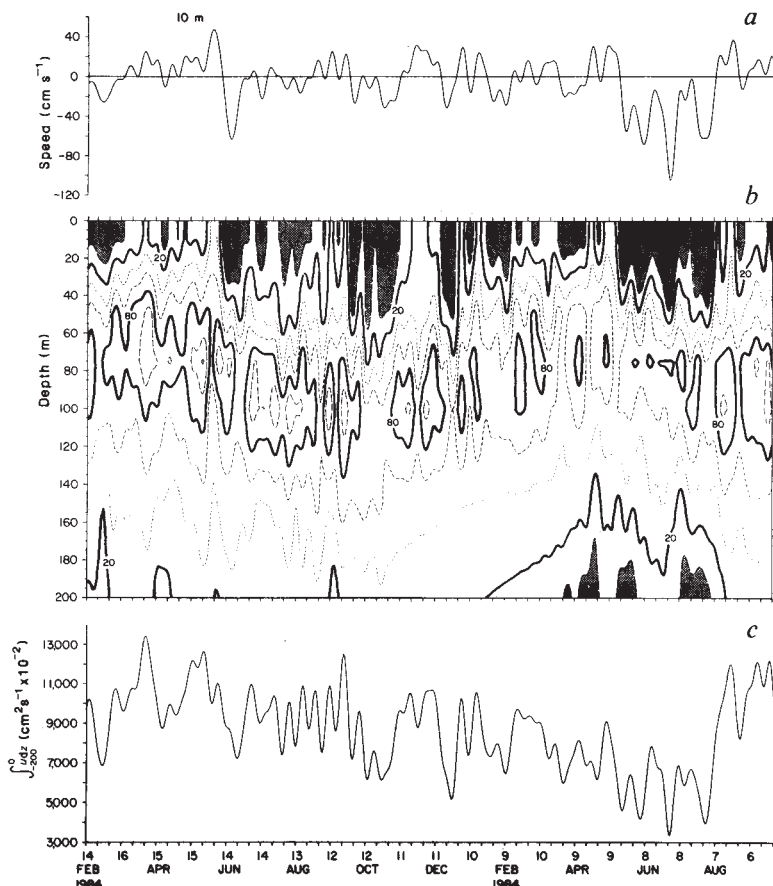
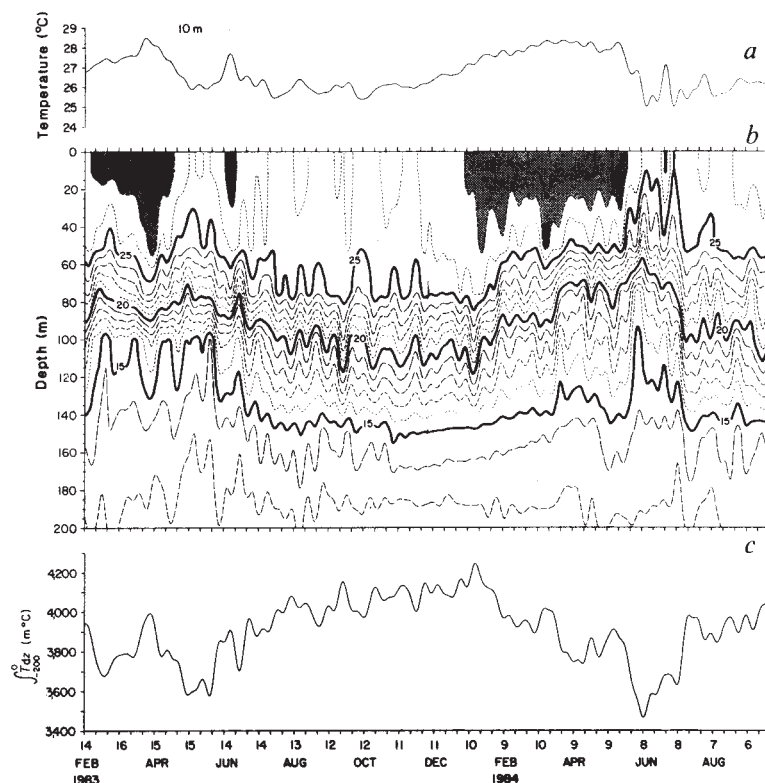


Fig. 2 Eastward component of velocity (u) observed over the upper 200 m on the Equator at 28° W from February 1983 to October 1984. *a*, u at 10 m depth; *b*, u isotach depths as a function of time; *c*, the vertical integral of u . The data have all been low-pass filtered to remove fluctuations with timescales <10 days. Regions of westward flow are stippled.

This is true for both the vertical position of a given isotherm within the thermocline as well as for SST. Not only does the Gulf of Guinea have the coldest SST along the Equator in the Atlantic Ocean during summer, it also has the warmest SST during winter. Owing to the increased duration of the response sequences the annual variations show a more sinusoidal appear-

ance at 4° W than at 28° W (where the thermocline attained an adjusted equilibrium to increased easterly wind stress for several months in 1983). Because it takes longer for the thermocline at 4° W to adjust to increased easterlies than at 28° W, SST is more nearly in phase with the thermocline at 4° W although SST still remains low even after the thermocline begins to deepen there.

Fig. 3 Thermal structure over the upper 110 m on the equator at 4° W from February 1983 to October 1984. *a*, The temperature at 10-m depth; *b*, isotherm depths as a function of time; *c*, vertically integrated temperature proportional to heat content. The data have all been low pass filtered to remove fluctuations with timescales < 10 days. Temperatures greater than 27 °C are stippled to highlight the warmest surface temperatures relative to the thermocline depth.

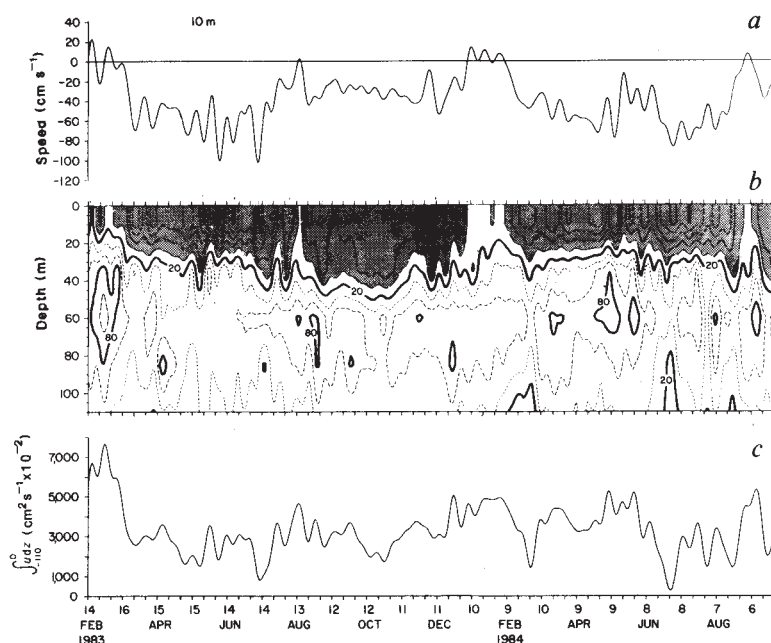
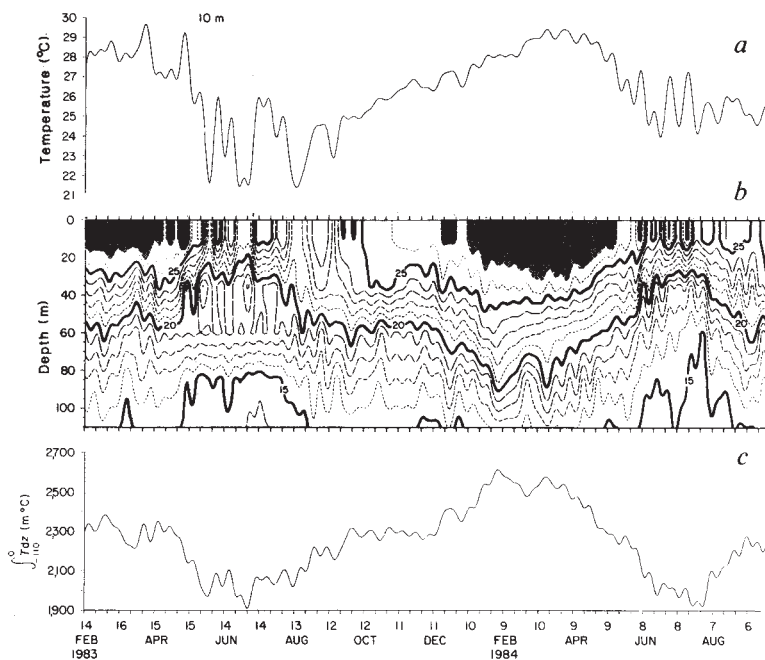


Fig. 4 Eastward component of velocity observed over the upper 110 m on the Equator at 4° W from February 1983 to October 1984. *a*, u at 10-m depth; *b*, u isotach depths as a function of time; *c*, the vertical integral of u . The data have all been low-pass filtered to remove fluctuations with timescales < 10 days. Regions of westward flow are stippled.

Second, we observe an additional shoaling period at 4° W between October and December 1983 which is not evident at 28° W. Third, the deepening response at 4° W beginning in mid-December 1983 is much more pronounced than that at 28° W resulting in a well-defined peak in vertically integrated temperature at 4° W during February and March 1984. The corresponding change in dynamic height between the surface and 110 m due to the increased temperature is ~8 dynamic cm in agreement with the FOCAL hydrographic section data obtained at that time⁸.

In the zonal velocity component, the principal difference between 4° W and 28° W is a much more persistent and stronger westward SEC on the surface at 4° W. Below the surface and down to the observational limit of 110 m the flow is eastward and the position of the EUC core deepens with the thermocline as at 28° W.

Before SEQUAL and FOCAL our view of the annual cycle was based on climatologically-averaged monthly fields derived

from decades of ship reports⁹⁻¹¹. Linear and non-linear numerical models¹²⁻¹⁴, driven by climatological wind fields, have been successful in emulating the vertically integrated climatological thermal field. The new basin-wide synoptic data show significant differences between the annual cycle of a given year and the climatological mean annual cycle. Rather than a slowly varying response nearly in phase with the easterly wind stress component the thermocline undergoes sequences of deepening and shoaling varying in both duration and magnitude along the Equator. These differences are related¹⁵ to the rapidly varying nature of the wind stress observed in any given year versus the slowly varying nature of the climatology; a property inherent in averaging data from many years with varying phase. Historical shipboard observed wind data when displayed by individual years¹⁶ rather than averaged show that rapid transitions are, indeed, the normal occurrence. Variations in the wind stress development from year to year will, therefore, give rise to interannual variability as a modulated annual cycle and this is evident in

comparing the 1984 and 1983 responses; years of relatively weak and strong surface easterlies respectively^{8,17}. At 28° W, the shoaling portion of the annual cycle shows the 20° C isotherm some 20 m higher in May and June 1984 than in 1983. At 4° W, the deepening portion of the cycle shows the 20° C isotherm some 30 m deeper in February and March 1984 than in 1983. The zonal distribution of upper ocean heat content was therefore different during these years with more heat in the eastern side of the basin during 1984.

SST along the Equator was also quite different. In 1984, from June to August, SST was roughly 1°C colder at 28° W (25°C versus 26°C) and 3°C warmer at 4° W (24.5°C versus 21.5°C) than over the same period in 1983. Consequently, during the cold season of 1984, the zonal SST gradient along the Equator had essentially vanished when it normally should have been a maximum. During the warm season (February and March) SST was warmer by roughly 1°C at both 28° W and 4° W in 1984 than in 1983. These SST differences seem to be linked with the differences reported in the position of the Intertropical Convergence Zone¹⁷ during these years.

While vertically integrated temperature and SST both differed from 1983 to 1984 it is reiterated that these two quantities are not related in a simple manner. At 28° W minimum SST coincides with a deepened thermocline. At 4° W minimum SST occurs while the thermocline is deepening. In either case, vertically integrated upper ocean temperature or heat content alone is not adequate to account for air-sea heat exchange since this process depends on SST.

Variations in the velocity field appear to be more complicated than those in the temperature field and the relative importance of horizontal and vertical advection in the heat balance has not yet been assessed. Note that at 28° W during March-June 1983 the transport per unit width on the Equator was nearly double the amount observed during those months in 1984.

We have presented a subset of the SEQUAL and FOCAL moored current meter data showing the evolution of the annual cycle in upper ocean zonal currents and temperature along the Equator. Differences found between climatological data inference and synoptic data realization are related to the rapidly varying nature of the easterly surface wind stress component observed in any given year versus the slowly varying climatology. In 1984, a year of relatively weak winds compared with 1983, the eastern portion of the basin showed increased values of vertically integrated temperature and SST during the winter/spring and summer seasons respectively. These changes in the zonal distribution of temperature along the Equator in the Atlantic Ocean have similarities with the El Niño occurrences in the Pacific Ocean.

This work was supported by the Oceanography Section, NSF under grant number OCE-8211848 and by ORSTOM, Paris. We thank T. Y. Tang and T. Weingartner for helpful discussions.

Oceanic conditions in the tropical Atlantic during 1983 and 1984

P. Hisard*, C. Henin*, R. Houghton†, B. Piton‡ & P. Rual‡

* ORSTOM/LPDA, Université Pierre et Marie Curie, 4 Place Jussieu, 75005 Paris, France

† Lamont-Doherty Geological Observatory of Columbia University, Palisades, New York, 10964, USA

‡ ORSTOM/COB, BP 337, Brest, France

During 1983 and 1984 oceanographers from France and the USA conducted an extensive field programme in the tropical Atlantic Ocean. The principal goal was to document the annual cycle, but unexpectedly a difference was observed between the two years. We describe here the observed variations in the thermal field and in the zonal currents along several meridians. In early 1984, the upper ocean especially on the eastern side of the ocean basin was substantially warmer than during the corresponding period a year earlier. Most of this change in the heat content was the result of a deeper thermocline. During this time an unusual eastward surface current appeared to the south of the Equator undoubtedly contributing to large zonal heat fluxes. The decrease in the surface winds¹ between 1983 and 1984 is presumably responsible for these changes. Inspection of the few available long records reveals that in the Atlantic warm events such as that of 1984 are rare and that the previous event with comparable amplitude occurred in 1963.

Oceanic conditions in the tropical Atlantic have a pronounced annual cycle which is almost in phase with the annual changes in the surface winds. When the winds near the Equator are relaxed, in March and April, horizontal thermal gradients are at a minimum and the surface currents are weak and westward. These currents feed the Brazilian coastal current which flows northward into the Gulf of Mexico. The intensification of the south-east trade winds from May onwards results in a dramatic change in the oceanic circulation. The westward surface flow, to the south of 3° N approximately, accelerates while an intense eastward surface current appears between 3° N and 9° N approximately². The Brazilian coastal current now veers offshore near 5° N to maintain the eastward flow³. The associated changes in the thermal field include a deepening of the isotherms in the western equatorial Atlantic, especially in the neighbourhood of 3° N, and an elevation of isotherms of the Gulf of Guinea and near 9° N in the western side of the ocean basin⁴. These changes in the thermal structure affect the sea-surface temperatures which decrease at, and to the south of, the Equator in the central and eastern Atlantic once the south-east trade winds start to strengthen.

The field work conducted in 1983 and 1984 was part of the Seasonal Response of the Equatorial Atlantic (SEQUAL) and Français Ocean et Climat dans l'Atlantique (FOCAL)

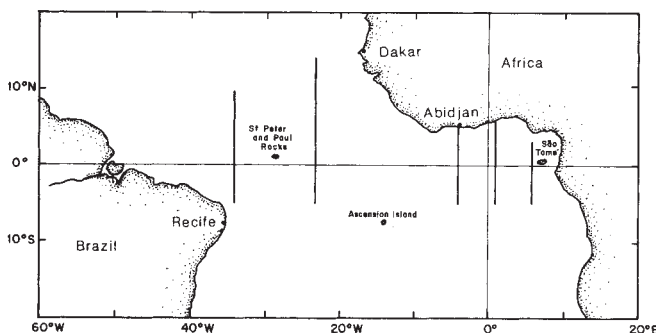


Fig. 1 Equatorial Atlantic Ocean with the meridional sections used in this study.

1. Bjerknes, J. *Tellus* **18**, 820-828 (1966).
2. Rasmusson, E. M. & Wallace, J. M. *Science* **222**, 1195-1202 (1983).
3. Moore, D. W. & Philander, S. G. H. in *The Sea Vol. 6* (eds Goldber, E. et al.) 319-361 (Wiley-Interscience, New York 1977).
4. Cane, M. & Sarachik, E. S. *J. mar. Res.* **35**, 395-432 (1977).
5. Garzoli, S. & Katz, E. J. *Geophys. Res. Lett.* **11**, 715-718 (1984).
6. Katz, E. J. *Geophys. Res. Lett.* **11**, 726-728 (1984).
7. Payne, R. E. *Geophys. Res. Lett.* **11**, 719-721 (1984).
8. Katz, E. J., Hisard, P., Verstraete, J. M. & Garzoli, S. L. *Nature* **322**, 245-247 (1986).
9. Hastenrath, S. & Lamb, P. J. *Climatic Atlas of the Tropical Atlantic and Eastern Pacific Ocean* (University of Wisconsin Press, 1977).
10. Hellerman, S. *Deep-Sea Res. Suppl.* **II** **26**, 63-75 (1979).
11. Merle, J. *Trav. Doc. ORSTOM* **82** (1978).
12. Busalacchi, A. J. & Picaut, J. *J. phys. Oceanogr.* **13**, 1564-1588 (1983).
13. Philander, S. G. H. & Pacanowski, R. C. *Geophys. Res. Lett.* **11**, 802-804 (1984).
14. du Penhoat, Y. & Treguier, A. M. *Geophys. Res. Lett.* **11**, 799-801 (1984).
15. Weisberg, R. H. & Tang, T. Y. *J. geophys. Res.* **90**, 7117-7128 (1985).
16. Servain, J., Picaut, J. & Merle, J. *J. Phys. Oceanogr.* **12**, 457-463 (1982).
17. Horel, J. D., Kousky, V. E. & Kagano, M. T. *Nature* **322**, 248-251 (1986).

Received 21 January; accepted 16 April 1986.

programmes, and it provided extensive observations of the thermal structure in the equatorial Atlantic during that period. Preliminary results from the first year (1983) measurements are given elsewhere⁵. Data for this study were derived from several sources. These include meridional hydrographic sections along 34° W, 23° W, 4° W, 1° E and 6° E (Fig. 1) that were made four times each year by the research vessels *Capricorne* and *Nizery*. On these sections, temperature and salinity were measured by means of a CTD (conductivity temperature/depth), currents were measured by means of a profiler⁶ at the station locations shown in Fig. 2. Additional data were derived from aircraft and ship-of-opportunity expendable bathythermographs.

We shall confine our discussion to the temperature structure along 23° W, 4° W and 6° E measured in January–February and in July–August of 1983 and 1984. The depth of the 20 °C isotherm along these sections is shown in Fig. 2. The 20 °C isotherm lies within the thermocline and is thus representative of the thermal structure in the upper 100 m of the water column. The most striking difference between the sections made in the two years is the increase in the average temperature of the upper ocean, between February 1983 and February 1984. This warming was most pronounced in the Gulf of Guinea where the thermocline in some places was 30 m deeper than normal in February 1984 throughout the entire section. The equatorial temperature profiles shown in Fig. 3 illustrate this change. In 1983, the thermocline depth was approximately equal to the climatological mean value⁷, whereas in 1984 it was approximately three standard deviations below the mean. This condition was so anomalous that in January and February 1984 the thermocline which usually slopes downward to the west was practically horizontal⁸. Time series measurement along the Equator indicate that the unusual warming started as early as November 1983 and persisted into May 1984.⁹ The warm water that accumulated in the Gulf of Guinea was unusually saline which indicates that its origin was the western side of the ocean basin.

Sections along 23° W show that the meridional gradient of shallow isotherms (not shown in Fig. 2) increased significantly to the south of the Equator between February 1983 and February 1984. This implies an eastward geostrophic current between the Equator and 5° S in February 1984. Normally the surface flow in this region is westward. Direct measurements with current profilers averaged over 0–20 m depth confirm this unusual eastward current which penetrated into the Gulf of Guinea (Fig. 4). The current is absent from sections made in April 1984 but appears again in July 1984. Presumably this current played an important role in transporting warm saline surface waters into the Gulf of Guinea thus increasing the average temperature of the upper ocean in that region. There is evidence that the eastward north equatorial countercurrent, to the north of 3° N, also contributed to the unusual accumulation of warm surface waters in the eastern equatorial Atlantic. Some of this warm water flowed southward along the African coast and may be a factor in inhibited coastal upwelling as far south as Angola and Namibia¹⁰.

The change in the surface winds^{1,8} between 1983 and 1984 was presumably responsible for the change in the oceanic circulation between those two years. The westward winds maintain the zonal density gradients in the ocean so that the unusually weak winds in early 1984 resulted in the reduction of the zonal slope of the thermocline during January and February and the associated eastward transfer of warm surface waters. That this transfer occurred primarily in the eastward current to the south of the Equator, and in the eastward north equatorial countercurrent to the north of 3° N approximately, suggests that there were significant changes in the curl of the wind during 1984. The curl is large in the neighbourhood of the Intertropical Convergence Zone (ITCZ) which is usually over the north equatorial countercurrent to the north of 3° N. It is intriguing that from satellite cloud images in early 1984 a second ITCZ was observed to the south of the Equator where the unusual eastward current

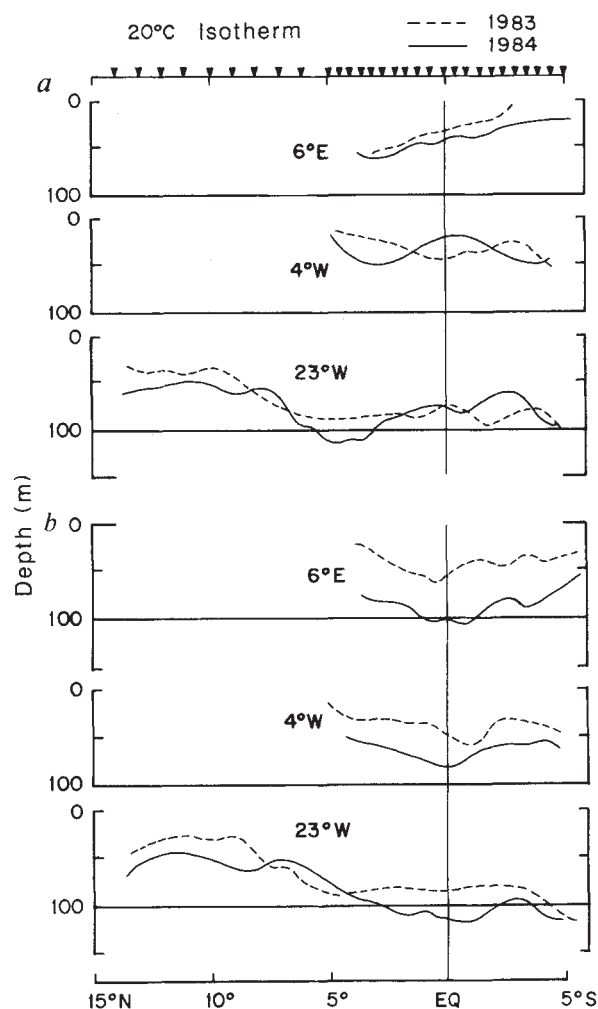


Fig. 2 Meridional sections of the depth of the 20 °C isotherm along 6° E, 4° W, and 23° W longitude during July–August (a) and January–February (b) for 1983 (dashed line) and 1984 (solid line). CTD station locations indicated at top.

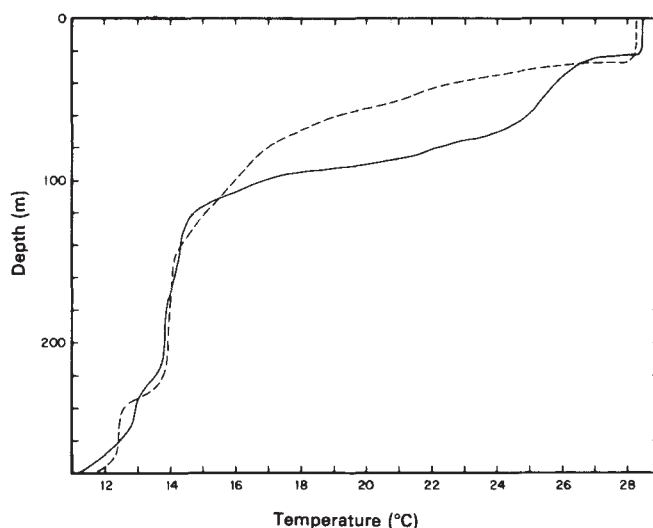


Fig. 3 Profiles of the temperature on the Equator at 4° W in February 1983 (dashed line) and 1984 (solid line).

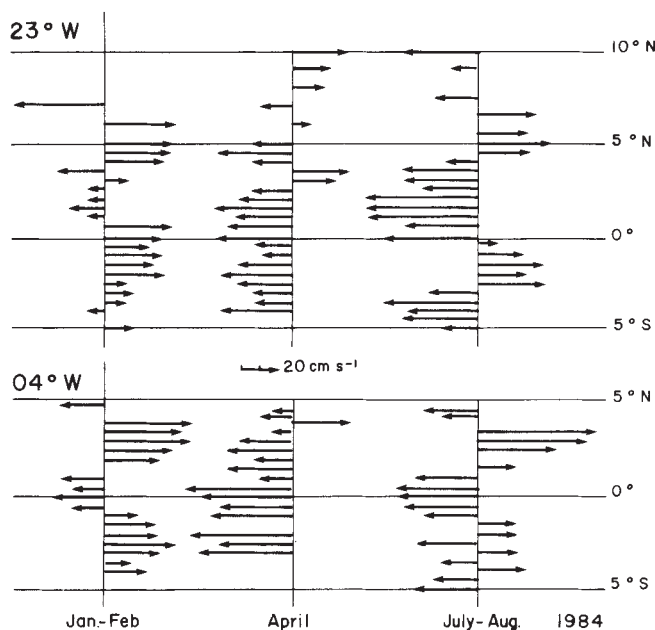


Fig. 4 The zonal component of the 0–20 m surface currents (arrows to the right are eastward) as measured along 23° W and 4° W in January–February, April and July–August 1984.

appeared (J. Citeau and B. Guillot, personal communication). The relations between this current, the ITCZ, and the curl of the wind are now being studied. Another topic under investigation is the relation between the north-westward flowing Brazilian coastal current and the eastward currents. This coastal current, as mentioned earlier, absorbs water from the westward currents and, in turn, maintains the eastward flow to the north for 3° N approximately. It is not yet known whether the unusual eastward current to the south of the Equator in early 1984 was fed by the Brazilian coastal current.

There are many similarities between oceanographic conditions in the tropical Atlantic in 1984 and 1963 when a comparable warm event occurred. Not only were sea-surface temperatures exceptionally high in 1963, but there were also reports of unusual eastward surface currents to the south of the Equator, and of the intrusion of warm saline water into the coastal regions of Angola and Namibia where local coastal upwelling was inhibited^{11–14}. Occurrences such as these are very similar to but much less frequent than El Niño events in the Pacific Ocean. In the Atlantic, interannual variability on shorter timescales has a relatively small amplitude but is, nonetheless, of considerable interest because the physical processes involved in the major events are likely to be the same as those involved with smaller amplitudes.

This work was supported in part by ORSTOM and by the NSF (grant OCE 81-17554). We thank S. G. H. Philander for assistance. This is Lamont-Doherty Geological Observatory contribution no. 3967.

Received 17 January; accepted 18 April 1984.

- Horel, J. D., Kousky, V. E. & Kagano, M. T. *Nature* **322**, 248–251 (1986).
- Katz, E. J. & Garzoli, S. L. *J. mar. Res.* **40**, 307–327 (1982).
- Bruce, J. G. & Kerling, J. L. *Geophys. Res. Lett.* **11**, 779–782 (1984).
- Merle, J. & Arnault, S. *J. mar. Res.* **43**, 267–288 (1985).
- Weisberg, R. W. (ed.) *Geophys. Res. Lett.* **11**, 713–803 (1984).
- Duing, W. & Johnson, D. *Deep Sea Res.* **19**, 259–276 (1972).
- Houghton, R. W. *J. phys. Oceanogr.* **13**, 2079–2081 (1983).
- Katz, E. J., Hisard, P., Verstraete, J. M. & Garzoli, S. L. *Nature* **322**, 245–247 (1986).
- Weisberg, R. & Collin, C. *Nature* **322**, 240–243 (1986).
- Shannon, L. V., Boyd, A. J., Brundrit, G. B. & Taunton-Clark, J. *J. mar. Res.* (in the press).
- Mazeika, P. A. *J. geophys. Res.* **73**, 5819–5828 (1968).
- Reid, J. L. *Nature* **203**, 182 (1964).
- Hisard, P. *Oceanol. Acta* **3**, 69–78 (1980).
- Voituriez, B. *Oceanogr. Tropic.* **18**, 185–199 (1983).

Annual change of sea surface slope along the Equator of the Atlantic Ocean in 1983 and 1984

Eli Joel Katz*, Philippe Hisard†, Jean-Marc Verstraete‡ & Silvia L. Garzoli*

* Lamont-Doherty Geological Observatory of Columbia University, Palisades, New York 10964, USA

† Université P.-M. Curie, 4 Place Jussieu, 75230 Paris, France

‡ O.R.S.T.O.M., 195 Rue Saint Jacques, 75005 Paris, France

For two years beginning in October 1982, oceanographic observations were made in the equatorial Atlantic to investigate the relationship between sea level, wind stress, sea surface slope (zonal pressure gradient, thermocline depth) and equatorial currents. Here we describe the sea surface slope derived from three complementary methods: hydrographic stations, pressure gauges and inverted echo sounders. The latter two have the advantage of yielding continuous time series, but depend on the hydrographic stations for an absolute reference. Together, the three provide a detailed description of the temporal variation of the sea surface slope which is then compared to a wind-stress time series. The dominant signal, in both sea slope and wind stress, is the annual cycle, although amplitude and phase vary interannually. The annual increase in sea surface slope along the Equator in the western and central basins lags the onset of the south-east trade wind. During the boreal winter of 1983–84 a strong rise in sea level occurred against the African coast, accompanied by a levelling of the sea surface to the west. At the same time, an almost complete relaxation of eastward wind stress on the Equator was observed near the centre of the basin.

The sea surface along much of the Equator in the Atlantic and Pacific oceans slopes upward to the west. This slope reflects a shallowing of the thermocline toward the east and results in a zonal pressure gradient in the uppermost 100 m which is essential to the existence of the Equatorial Undercurrent¹. The

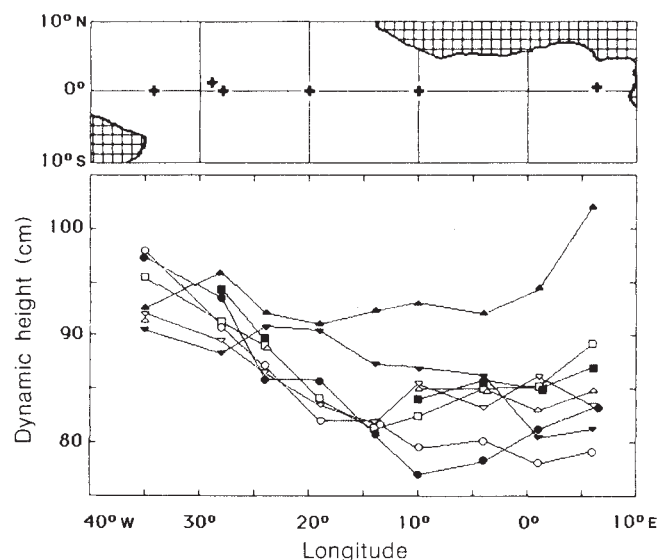


Fig. 1 Quarter-annual dynamic height of the sea surface, relative to 500 m. □ (■), October–November 1982 (1983); △ (▲), January–February 1983 (1984); ▽ (▼), April–May 1983 (1984); ○ (●), July–August 1983 (1984). Individual points are either meridional averages of casts within 0.5° of the Equator or zonal averages of casts within a 5° band along the Equator. Locations of inverted echo sounders, a wind recorder and pressure gauges discussed in the text are indicated in the chart above. The shaded areas are the bounding continents.

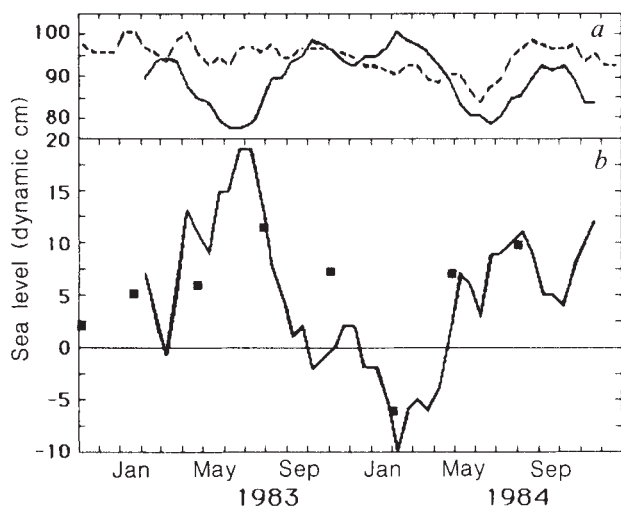


Fig. 2 Sea level difference (b) between St Peter and St Paul Rocks (29° W) and São Tomé island (7° E). The individual traces are shown in a (dashed line, 29° W; solid line, 7° E). The dynamic level at each site is determined by comparison with hydrocast dynamic height, 0 relative to 500 m. The differences in dynamic height from nearby hydrographic stations (near the Equator at 28° W and 6° E) are shown as points, although the casts were not simultaneous.

latter is a major zonal current in the tropics which, in 1979, transported an annual average of $20 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ of near-surface water from west to east in the Atlantic^{2,3}. A change in the mean level of the sea surface is a direct indication of a change in heat storage in the upper few hundred metres of that region, and a zonally differential change in mean level varies the acceleration force imposed on the fluid by the zonal pressure gradient.

The pressure gradient has a seasonal signal: weak in the boreal spring followed by a stronger gradient in the summer and autumn⁴⁻⁶. Six quasi-synoptic observations, obtained during three different years, found gradients at the surface ranging from 0.7 to $7.3 \times 10^{-5} \text{ dyn g}^{-1}$ (ref. 4). (The values are reported as accelerations obtained by dividing the pressure gradient by a density of $\sim 1 \text{ g cm}^{-3}$). Furthermore, the seasonal cycle is roughly in phase with the zonal wind field, both as reported by the ships during the discrete hydrographic observations and as compiled from the historical record⁴. Incomplete observations during 1979 of wind, sea surface height in the west and transport of the equatorial currents were compared to a simplified wind-forced model with encouraging results⁷.

Hydrographic data were obtained during eight cruises of both the *Capricorne* and *Nizery* between October 1982 and August 1984. Figure 1 shows the result of this work as the dynamic height at the surface, relative to 500 m. Gradients derived from the data west of the Gulf of Guinea are shown in a later figure. The slope of the sea surface in the Gulf is generally weak and without a preferred direction, except in January–February 1984, when there was a rapid rise in height approaching the African coast.

Pressure gauge data (Fig. 2) were obtained at 7° E and 29° W. The sensors were placed at depths of $<10 \text{ m}$ (in order to be equivalent to tide gauges), adjacent to the only two equatorial islands in the Atlantic; namely, São Tomé (0°15' N, 6°40' E) and St Peter and St Paul Rocks (0°55' N, 29°21' W). Half-monthly averages were compared with the average of several hydrocasts, each made seven times at each site during the record in order to get a consistent dynamic level between the two sites. The standard deviation of the difference between hydrographic and pressure gauge data, after adjustment for the mean, was $\pm 3 \text{ cm}$. The difference between the two gauges is plotted in Fig. 2b.

As shown in Fig. 1, the difference in dynamic height between 7° E and 29° W combines both the westward upward slope and the Gulf of Guinea slope. Thus the negative (westward downward) slope shown in Fig. 2 from December 1983 to mid-April 1984 is not a basin-wide reversal in surface tilt, but probably a neutral slope over much of the basin and a large height increase against the African coast. It confirms the suggestion from the hydrographic data that this condition did not occur in the previous 12-month period. The second clear contrast between the two years is the size of the positive difference in the summer, which was only half as large in 1984. This is not seen in the hydrographic data because of the discrete sampling.

Of the six inverted echo sounders deployed along the Equator during 1983–84, results from the four at 34° W, 28° W, 20° W and 10° W are discussed here. (A record at 1° W showed no substantial difference from that at 10° W and a short record at 38.5° W is omitted.) An inverted echo sounder is similar to a tide gauge in that it provides a measure of the vertically integrated temperature of the entire water column. To convert this to dynamic height, the hydrocasts are used to determine mean values, and historical data⁷ are used to relate a change in the observed travel time to a change in dynamic height. The standard deviation of the difference between the sounder-derived dynamic height and the hydrographic data is ± 2 dynamic cm.

The four resulting time series were low-pass filtered to remove tides and inertia-gravity waves. A linear regression analysis was then performed zonally on the data once every 10 days: the result is shown in Fig. 3, along with an indication of the goodness of fit. Also shown is the pressure gradient estimated from the hydrocast data (Fig. 1) for the same region. The 17-month time series derived from the four sounders shows the development of the zonal pressure gradient in the boreal spring of 1983, its gradual relaxation over a 9-month period and its subsequent redevelopment in 1984. This redevelopment occurred very rapidly during the first 20 days of June; the 1984 maximum exceeded that of 1983 by $\sim 20\%$.

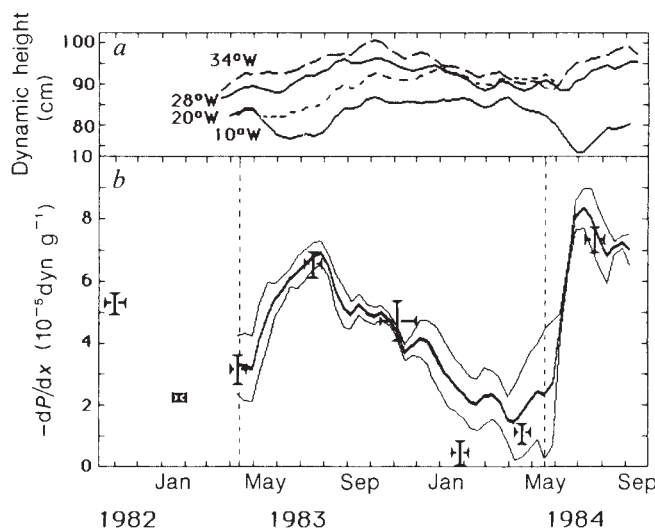


Fig. 3 Dynamic height (a) and zonal pressure gradient (b) between 34° W and 10° W along the equator. The solid line in b is the gradient computed at 10-day intervals from four inverted echo sounder records (shown in a). The bounding curves are determined from the standard error of the slope of the linear regression and, while not a true measure of statistical significance, they are a measure of linearity. The points (with error bars) are from the hydrocast data (Fig. 1) for the same region. The horizontal bar shows the time interval during which the casts were made; the vertical bar indicates the standard error of the slope (from a linear regression of (typically) six averages as plotted in Fig. 1). The dashed lines indicate the time of onset, each year, of the south-east trade wind (see Fig. 4).

In general, the hydrocast data (points in Fig. 3) yielded the same gradients, but with one important discrepancy. In the latter half of January 1984, the hydrographic data show no significant gradient. From Fig. 1 we note that this section (solid triangles, apex up) recorded an unusual sea surface slope, and a higher surface at every site (except for the westernmost location) than at any other time in the two years. Also, the 1984 surface is typically 10 cm higher than at the comparable time in 1983, increasing to nearly 20 cm near the eastern boundary. This anomaly is indirectly confirmed by the pressure gauge data (Fig. 2), which further suggests that this situation, although most pronounced near the time of the survey, was sustained during the months of January–March.

In contrast, the sounders did not indicate a minimum in pressure gradient during these months. Neither the sounder at 10° W, nor the one at 1° W not included in Fig. 3, indicated a high surface height. Indeed, the comparison between sounders and hydrocasts at these two sites showed discrepancies of 6 and 8 cm, several standard deviations more than the general result from the comparison. We cannot yet explain the different results from the two observational methods.

The objective of obtaining a time series of the zonal pressure gradient is to compare it with the synoptic wind field. The annual cycle of the trade winds along the Equator and west of the Gulf of Guinea results from the meridional migration of the Inter-tropical Convergence Zone (ITCZ), which is located near the Equator at the beginning of the year and reaches to beyond 10° N six months later. Horel *et al.*⁸ remark on a rapid relaxation of the near-surface wind field between November 1983 and January 1984, during which time the ITCZ shifted southward and then remained south of the Equator until May. Unlike the oceanographic data, which have only a meager and patchy historical base, the meteorological data are sufficiently complete for Horel *et al.*⁸ to conclude that conditions over the Atlantic in 1983–84 constituted an extreme event.

Our interest here is in the surface wind field directly over the Equator, in the band from the Gulf of Guinea to the South American coast. To characterize that wind field a wind recorder was placed on St Peter and St Paul Rocks: Fig. 4 shows the zonal wind stress computed from this record. From the historical data we know that this location is typical of the central and western reaches of the equatorial Atlantic, with stress increasing from east to west⁹. The record in Fig. 4 shows sudden increases in westward wind stress in the boreal spring of each year. A similar picture was obtained at the same site in 1979¹⁰ and in 1985 (S.L.G. and E.J.K., unpublished data), but the date of onset varies each year over a range of ~2 months.

The low wind stress observed from mid-January to May 1984 reflects the southerly location of the ITCZ in early 1984. The negative (eastward) stress in May is unusual. In an objective analysis of ship-reported winds (1964–84), Jacques Servain (personal communication) computed a monthly average eastward wind stress in the area of St Peter and St Paul Rocks only for May and June 1968. In 1983 and 1984, the onset of high wind stress leading to the seasonal trade wind occurred on 5 April and 19 May, respectively. These dates are indicated by dashed lines in Fig. 3, and they show the near simultaneity of the westward wind onset and the build-up of the zonal pressure gradient. A lag-correlation calculation between the data of Figs 3 and 4 indicates that the ocean surface height lagged the wind (at this one site) by 10–15 days. We note that the lower wind stress in mid-1984, relative to 1983, is associated with a larger maximum zonal pressure gradient in 1984.

The quarter-annual hydrographic surveys provide the densest set of classical observations yet obtained in the equatorial Atlantic, and the inverted echo sounder array is unique. The data confirm our previous understanding that the basin comprises two different regions of response, roughly delineated by the Gulf of Guinea. To the west of the gulf the hydrographic sampling proved adequate to define the annual variation of the

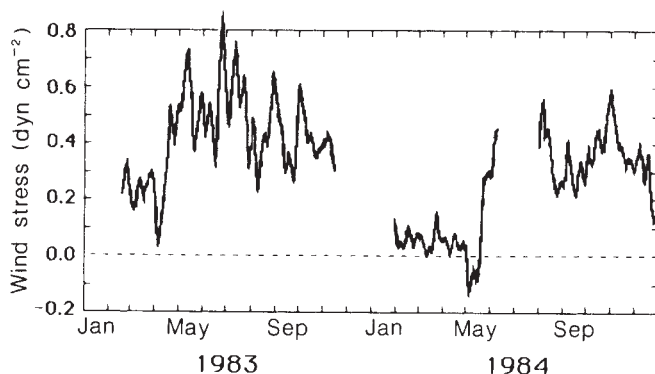


Fig. 4 Low-pass-filtered westward wind stress at St Peter and St Paul Rocks (29° W), 1983–84. The filter admits $>95\%$ of the variance for frequencies less than 0.1 cycle per day.

zonal pressure gradient, relative to the results from the continuously sampling sounders. The range of values, 0.8×10^{-5} dyn g^{-1} , is comparable to earlier estimates. The previous low value, obtained in the spring of 1963, was also associated with a particularly strong relaxation of the trade winds⁴. Both methods of observation show a slow and nearly monotonic decrease in zonal pressure gradient from August 1983 until May 1984, followed by a sharper and larger increase in gradient in the spring of 1984, relative to 1983. The hydrocast data show an anomalously high surface (deep thermocline) in January–February 1984 not evident in the sounder data: this difference needs further investigation.

The pressure gauge data (Fig. 2) compare the western central region with the eastern boundary of the Gulf of Guinea and also show an annual signal superimposed on an interannual difference in level. Around July of each year, a peak is seen in the sea surface height difference between west and east, and a minimum is found to occur around February. In 1984 the minimum is strongly negative (that is, the thermocline is deeper in the east), in agreement with the hydrocast data.

The zonal wind stress from 29° W (Fig. 4) has a boxcar-like time dependence: low in the boreal spring when the ITCZ is at or near the Equator and high when the south-east trade winds prevail. There is however an interannual difference, with the amplitude of the zonal stress in 1984 consistently lower than that in 1983 by ~ 0.15 dyn cm^{-2} throughout the entire year. The lower stress during 1984, however, resulted in a larger zonal pressure gradient west of the Gulf of Guinea which, unless opposed by other differences, would be expected to alter the equatorial circulation. The timing of the onset of the south-east trade winds is closely correlated with the build-up of the zonal pressure gradient. Our data imply a response time for the ocean of 10–15 days.

We acknowledge the financial support of the NSF (grants OCE 81-17000 and OCE 82-09892) and O.R.S.T.O.M. C. Henin and B. Piton collaborated with P.H. in obtaining the hydrographic data; G. Philander made helpful comments on an early draft. Lamont–Doherty Geological Observatory contribution no. 3966.

Received 15 January; accepted 27 March 1986.

1. Fofonoff, N. P. & Montgomery, R. B. *Tellus* 7, 518–521 (1955).
2. Katz, E. J. *et al. Oceanol. Acta* 4, 445–450 (1981).
3. Burkov, V. A., Zubin, A. B., Titov, V. B. & Kharlamov, A. I. *Soviet Met. Hydrol.* 9, 81–88 (1981).
4. Katz, E. J. *et al. J. mar. Res.* 35, 293–307 (1977).
5. Merle, J. *Atlas Hydrologique Saisonnier de l'Océan Atlantique Intertropical (Travaux et Documents de l'O.R.S.T.O.M. No. 82)* (Paris, 1978).
6. Lass, H. U. *et al. Oceanol. Acta* 6, 3–11 (1983).
7. Katz, E. J. & Garzoli, S. *J. mar. Res.* 40, 307–327 (1982).
8. Horel, J. D., Kousky, V. E. & Kagano, M. T. *Nature* 322, 248–251 (1986).
9. Hastenrath, S. & Lamb, P. J. *Climate Atlas of the Tropical Atlantic and Eastern Pacific Oceans* (University of Wisconsin Press, Madison, 1977).
10. Garzoli, S. L., Katz, E. J., Panitz, H. J. & Speth, P. *Oceanol. Acta* 5, 281–288 (1982).

Atmospheric conditions in the Atlantic sector during 1983 and 1984

John D. Horel*, Vernon E. Kousky†
& Mary T. Kagano‡

* Climate Research Group, Scripps Institution of Oceanography,
La Jolla, California 92093, USA

† Climate Analysis Center, National Meteorological Center,
NWS/NOAA, Washington DC 20233, USA

‡ Instituto de Pesquisas Espaciais, 12.200 Sao Jose dos Campos,
SP, Brazil

The dramatic global climate variability during the period from 1982 to 1984 has presented a unique opportunity to diagnose and interpret the role of air/sea interactions over the tropical oceans. Recent reviews have summarized the observational and numerical evidence for air/sea interactions over the equatorial Pacific Ocean associated with the El Niño/Southern Oscillation (ENSO) phenomenon¹⁻³. Here we present details of the atmospheric conditions in the Atlantic sector between 1983 and 1984. During the latter half of 1983, the atmospheric circulation over the equatorial Atlantic underwent a dramatic reversal: surface trade winds were substantially reduced; surface pressures decreased; and cloudiness and rainfall increased over the ocean and adjacent regions of north-east Brazil. These regional changes over the equatorial Atlantic coincided with planetary-scale adjustments in the tropical atmosphere. While the atmospheric circulation over the tropical Pacific was experiencing an extraordinary departure from normal during 1983, as evidenced by a reversal in direction of the surface winds, the circulation over the Atlantic was in a building-up phase with stronger than usual surface winds. As the atmospheric circulation over the Pacific returned to normal during the latter half of 1983, the trade winds over the Atlantic relaxed.

To place the atmospheric conditions over the Atlantic sector during 1983-84 in context, it is helpful to review the salient features of the global tropical circulation during this period. More detailed accounts of the month-to-month variability of the global circulation are given in the seasonal reviews⁴⁻¹⁰ and monthly *Climate Diagnostics Bulletins*. Over the equatorial Pacific, the largest climatic anomaly of the atmosphere-ocean system during the past 100 years was observed during the latter half of 1982 and early part of 1983 (refs 1-3). The conditions during this period represented an extreme 'warm' phase of the ENSO phenomenon, where 'warm' refers to the unprecedented increase in sea surface temperature in the equatorial Pacific. Some of the significant atmospheric circulation features observed during this period are: weakening of the Pacific trade winds and strengthening of the Atlantic trade winds; reduction of surface pressures over the eastern equatorial Pacific and increased pressures over Indonesia and the equatorial Atlantic; torrential rainfall over the central and eastern equatorial Pacific and drought over Indonesia and north-east Brazil.

While the 1982-83 ENSO warm phase in the equatorial Pacific has been well documented, its collapse during mid-1983 and

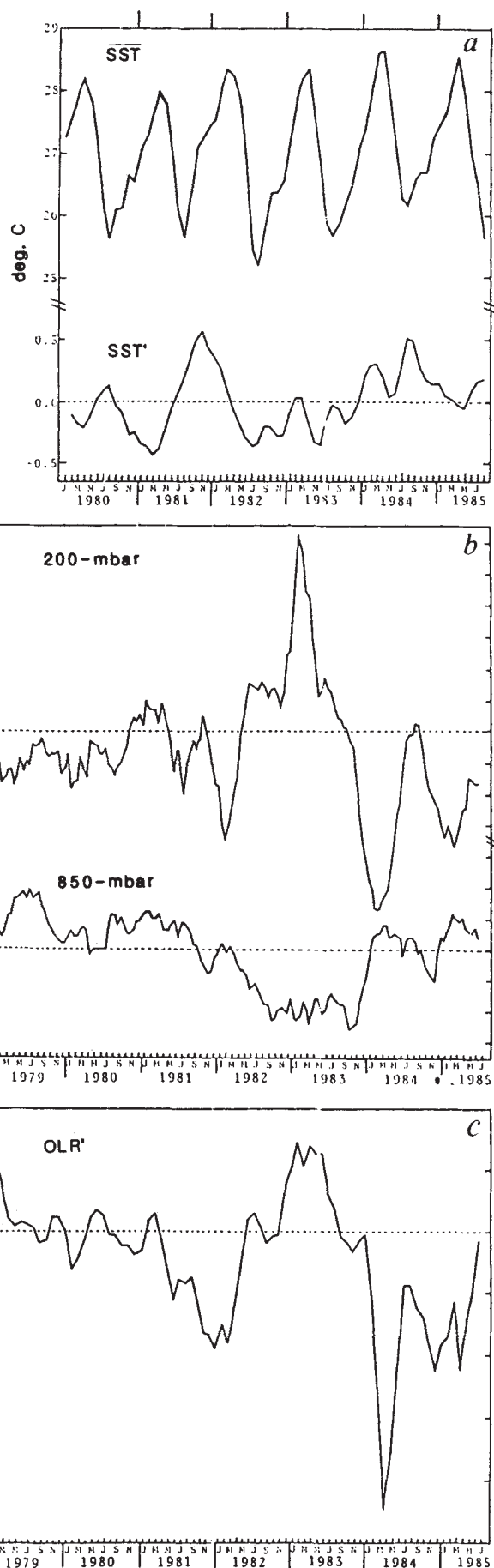


Fig. 1 Indices of: *a*, monthly averaged sea surface temperature (top) and departures (bottom) of the monthly averages from the climatological means¹⁸; *b*, 200-mbar zonal wind anomalies (top) and 850-mbar zonal wind anomalies (bottom); *c*, monthly anomalies of outgoing long-wave radiation. The indices are based on observations within the region 5° N to 5° S and from 35° W to 10° W. The anomalies of 850- and 200-mbar wind (outgoing long wave radiation) are computed with respect to the 1978-83 (1974-83) base period. To smooth month-to-month irregularities in the data, a 3-month running mean filter has been applied to the indices. The 1980-85 mean anomaly has been removed from the sea surface temperature anomaly index, to compensate for trends in the long-term sea surface temperature record.

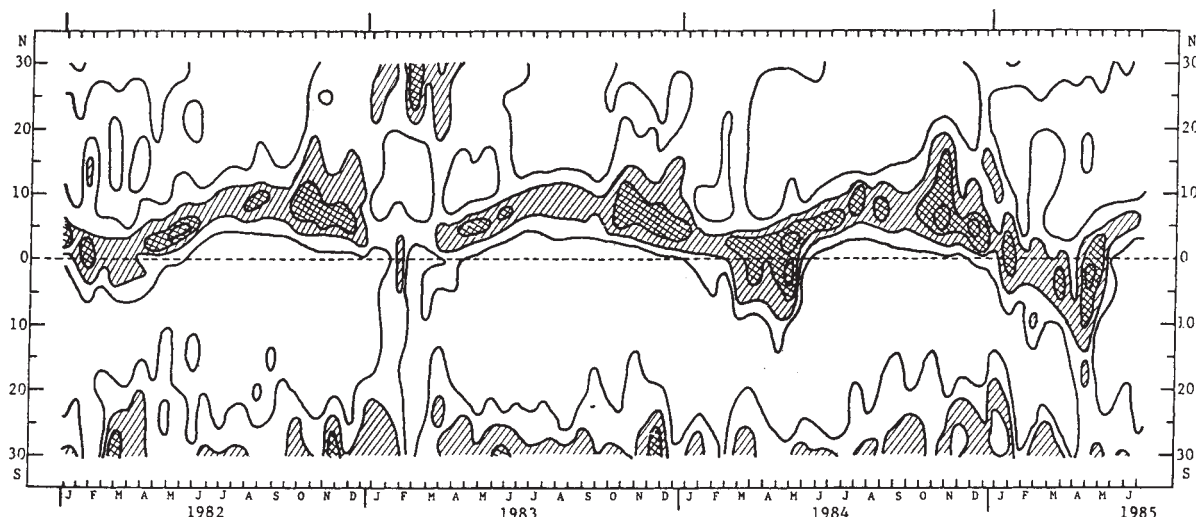


Fig. 2 Time versus latitude plot of outgoing long-wave radiation (OLR) averaged over the latitude band 20°–35° W. OLR values $>240 \text{ W m}^{-2}$ are indicated by a lack of hatching; hatching denotes OLR values between 220 and 240 W m^{-2} ; cross-hatching indicates values less than 220 W m^{-2} ; the contour interval is 20 W m^{-2} .

subsequent reversal into a 'cold' ENSO phase have received less attention. In many respects, the tropical circulation during 1984 was as unusual as that during the preceding year: decreased ocean temperatures in the central equatorial Pacific and increased temperatures in the equatorial Atlantic; strengthening of the Pacific trade winds and weakening of the Atlantic trade winds; higher pressures over the eastern equatorial Pacific and decreased pressures over Indonesia and the equatorial Atlantic; dry conditions over the central equatorial Pacific and above normal rainfall over Indonesia and north-east Brazil.

Briefly, the anomalous atmospheric conditions during 1983 (1984) over the equatorial Pacific can be explained largely in terms of the atmospheric response to the unusually warm (cold) ocean surface^{1–3,11–13}. Similar processes undoubtedly had a role over the equatorial Atlantic during 1984. Because of the sensitivity of the tropical atmosphere to relatively small changes in equatorial sea surface temperature, sea surface temperature is often regarded more as an atmospheric variable than an oceanic one. In the following, we will contrast the changes in sea surface temperature in the equatorial Atlantic to those in the overlying atmosphere during 1983–84. In addition, we will compare the relative magnitudes and durations of the anomalous conditions in the Atlantic sector during this period to those over the Pacific associated with ENSO.

Figure 1 contrasts a number of indices of the sea surface temperature and selected atmospheric fields. The month-to-month changes in sea surface temperature averaged over the central equatorial Atlantic from 5° N to 5° S and 35° W to 10° W are shown in Fig. 1a for the period 1980–85. During every year, sea surface temperature exhibits a pronounced annual cycle with the highest (lowest) temperatures during February–April (August–October)^{14,15}. Note that the warmest ocean surface ($>28.5^\circ\text{C}$) was evident during February–May 1984. In contrast, averages of sea surface temperature for the central equatorial Pacific exhibit less of an annual cycle and more variability from year to year^{16,17}.

The departures from climatology¹⁸ of central equatorial Atlantic sea surface temperature are also shown in Fig. 1a. Large positive anomalies were observed in late 1981; however, these occurred primarily during the cool phase of the annual cycle. Sea surface temperature was near or slightly below normal from mid-1982 until late 1983. After the beginning of 1984, the sea surface became anomalously warm in the equatorial Atlantic, with peak anomalies occurring during August–September. The anomalies decreased thereafter but remained positive until June 1985. The magnitude ($<1^\circ\text{C}$) of the sea surface temperature

increase in the equatorial Atlantic during 1984 is modest compared with that in the Pacific during 1983 ($2\text{--}3^\circ\text{C}$).

To first order, the fluctuations in the atmospheric circulation over the equatorial Atlantic can be viewed as perturbations of the mean zonally oriented circulation cell characterized by rising motion over the Amazon Basin, surface inflow from the Atlantic trade winds (directed from east to west), and outflow aloft (directed from west to east). Superimposed on this mean zonal circulation are large seasonal meridional migrations of the Inter-tropical Convergence Zone (ITCZ), an east–west oriented band of extensive cloudiness and rainfall. Indices of 850-mbar zonal wind (indicative of the surface trade-wind field) and 200-mbar zonal wind (representative of the flow in the upper troposphere) for the region used for the sea surface temperature index are shown in Fig. 1b. Positive (negative) values of the 850-mbar wind index indicate weaker (stronger) than normal surface trade winds over the equatorial Atlantic. The most striking features of the near-surface wind field are the gradual buildup of the trade winds during mid-1982 and their rapid relaxation between November 1983 and January 1984. The effects of these changes in surface wind stress on the ocean circulation are discussed in refs 19–24.

The out-of-phase relationship between near-surface and upper tropospheric winds is evident in Fig. 1b. In contrast to the near-surface winds, year-to-year fluctuations in the upper tropospheric winds are strongly modulated by the annual cycle such that the largest anomalies are observed from January to April of 1983 and 1984.

Figure 1c shows the month-to-month changes in outgoing long-wave radiation averaged over the region used for the other indices. Outgoing long-wave radiation provides an indication of cloudiness and deep tropical cumulus convection: positive (negative) anomalies of outgoing long-wave radiation indicate warmer (colder) cloud-top temperatures and less (more) cloudiness and rainfall than usual. Rainfall and cloudiness were inhibited during February–May 1983 while they increased substantially during February–May 1984.

As mentioned above, variations in cloudiness and rainfall over the equatorial Atlantic are strongly modulated by the seasonal meridional migrations of the ITCZ. The ITCZ normally extends furthest south and straddles the Equator in the western and central Atlantic from February to April. Figure 2 shows the meridional migrations of the ITCZ during the period from 1982 to 1985 in terms of a time–latitude plot of outgoing long-wave radiation averaged over the longitude band 20° W to 35° W. Note that the ITCZ was noticeably weak and slightly further north

during February–April 1983, compared with the other years. During March–May of 1984 and 1985, the ITCZ was vigorous and expanded southward.

The variations in cloudiness and rainfall evident in Figs 1c and 2 for the equatorial Atlantic extend westward into the region of north-east Brazil. Figure 3 contrasts the distribution of anomalous rainfall during 1983 and 1984 during the normally rainy seasons of March–May. The anomalies of rainfall at each station have been normalized by the climatological mean rainfall for that season so that the figures are not biased towards the Amazon regions, which receive more rainfall than the drought-prone areas of north-east Brazil. During 1983, most stations in north-east Brazil measured less than half of their normal rainfall amounts for that season, whereas the opposite was true during 1984.

Half-monthly values of sea-level pressure were substantially lower over the entire South Atlantic during 1984 in comparison with corresponding periods during 1982–83 (not shown). The reduction in surface pressure is consistent with the decreased strength of the trade winds. Sea-level pressure in the eastern tropical Atlantic returned to pre-1984 levels by September–November 1984. However, in the western tropical Atlantic, sea-level pressure remained lower than that during 1982–83 until June 1985.

The temporal and spatial evolution of the variations in ocean surface temperature and atmospheric circulation can be summarized as follows. Beginning in early 1984, sea surface temperature began to rise sharply across the entire equatorial strip, particularly to the south of the Equator. The warmest temperatures were observed along the Equator from March to May 1984, the season during which the equatorial Atlantic is normally warmest. Anomalies of roughly 1°C in sea surface temperature persisted through the remainder of 1984 to the south of the Equator, while temperatures were normal or slightly below normal to the north of 5°N . Although temperatures in the eastern equatorial ocean returned to near normal by early 1985, above-normal temperatures remained in the western Atlantic until June 1985.

The atmospheric circulation over the tropical Atlantic during early 1983 was in a highly perturbed state, with higher pressures than normal, stronger surface trade winds, a weakened and northward-displaced ITCZ, and drought over north-east Brazil. As the surface temperatures in the equatorial Atlantic were close to normal during this period, it is unlikely that this anomalous atmospheric circulation was forced from below. Concurrent with the anomalous circulation over the Atlantic, an extreme warm phase of ENSO was observed over the equatorial Pacific. As the low-frequency fluctuations of the tropical atmosphere are highly coupled over long distances, it is plausible that the anomalous conditions over the Atlantic were a remote readjustment of the tropical circulation to the unusually intense convection and rising motion over the eastern equatorial Pacific. Possible links between the warm phase of ENSO in the Pacific and weakening of the Atlantic ITCZ and drought in north-east Brazil have been discussed elsewhere²⁵.

During the period from November 1983 to January 1984, the surface trade-wind and surface pressure fields weakened. The reduction in surface wind stress led to alterations in the circulation and thermal fields of the equatorial Atlantic. As sea surface temperatures rose during early 1984, the Atlantic ITCZ shifted southwards and remained near and to the south of the Equator in the vicinity of the warmest sea surface temperature ($>27^{\circ}\text{C}$) until May. The anomalous pattern of near-normal sea surface temperatures to the south of the Equator and below-normal temperatures to the north of the Equator has been linked before to abundant rainfall in north-east Brazil and drought over the sub-Saharan regions of Africa^{26,27}.

Although surface pressures and surface trade winds increased perceptibly over the eastern Atlantic and north of the Equator towards the end of 1984, the ITCZ again migrated south of the

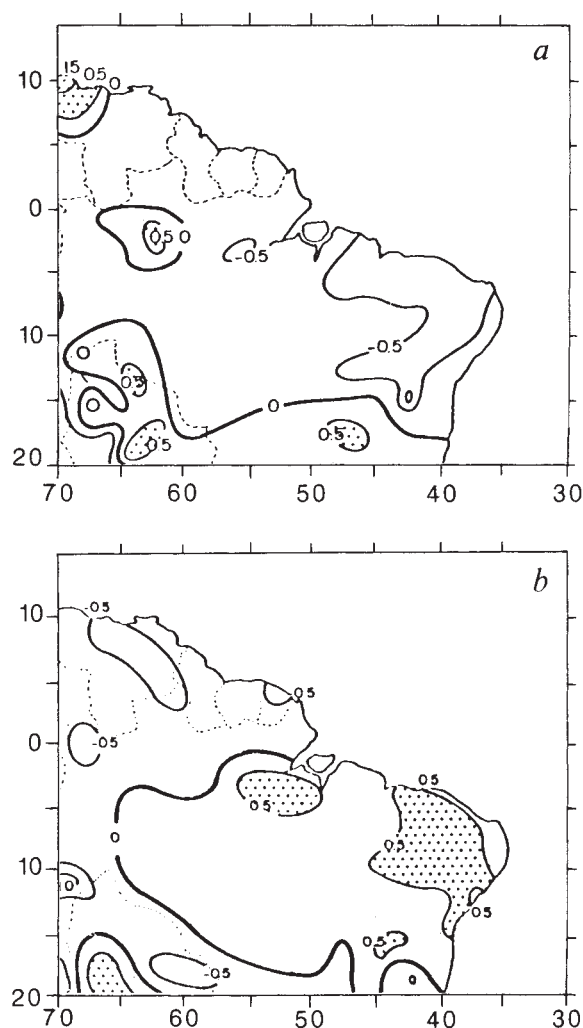


Fig. 3 Normalized precipitation deviations over north-east Brazil during a, March–May 1983 and b, March–May 1984. Values are calculated by determining the precipitation deviations from the mean and then dividing by the mean. The contour interval (in non-dimensional units) is 0.5; stippling indicates regions where seasonal rainfall totals were >0.5 of the seasonal mean.

Equator during February–May 1985 in the western Atlantic. This led to another season with above-normal rainfall in north-east Brazil. Sea surface temperatures gradually returned to near normal by the end of May 1985 in the equatorial Atlantic, and rainfall in sub-Saharan regions of Africa during the 1985 rainy season was much more plentiful than that during the previous two years.

The atmospheric circulation observed during the warm 1984 Atlantic event has many similarities to the warm 1983 ENSO event over the Pacific. The weakened trade winds, southward displacement of the ITCZ and enhanced equatorial convection are also features common to warm events in both the Atlantic and Pacific regions. However, several features are distinctly different. A major component of the warm Pacific event is the east–west shift of convection from the maritime continent towards the dateline and ‘see-saw’ of surface pressure, with lower pressures in the southeastern Pacific and higher pressures over the maritime continent^{1–3}. In contrast, the observed changes in cloudiness over the Atlantic were oriented primarily meridionally; convective activity over the Amazon basin exhibited little east–west movement during 1984 and pressures were reduced nearly uniformly across the southern Atlantic. Furthermore, the magnitude of the circulation changes in the Atlantic sector are

considerably weaker (roughly one half to one third) than those during the 1983 Pacific warm event and even slightly weaker than those during other such Pacific warm events, for example, those of 1972 and 1957.

The development of an unusually warm equatorial Atlantic, relaxed trade winds, and southward-displaced ITCZ are consistent with current theories on air/sea interaction over equatorial oceans. However, the conditions over the Atlantic during 1983–84 constitute an extreme event, much like the extreme nature of the 1982–83 warm event in the Pacific.

We thank Kathy Stevenson and John Knopman for assistance. Partial support for J.D.H. was provided by the Climate Dynamics Program under NSF grant ATM 84-04548. These results were presented at the CCCO Tropical Atlantic Panel Meeting held in Rio de Janeiro, Brazil 9–13 September 1985. The *Climate Diagnostics Bulletin* is prepared by staff of the Climate Analysis Center, National Weather Service, Washington DC.

Received 30 January; accepted 10 April 1986.

- Philander, S. G. H. *Nature* **302**, 295–301 (1983).
- Gill, A. E. & Rasmusson, E. M. *Nature* **306**, 226–234 (1983).
- Rasmusson, E. M. & Wallace, J. M. *Science* **222**, 1195–1202 (1983).
- Quiroz, R. S. *Mon. Weath. Rev.* **111**, 1685–1706 (1983).
- Chen, W. Y. *Mon. Weath. Rev.* **111**, 2371–2384 (1983).
- Ropelewski, C. F. *Mon. Weath. Rev.* **112**, 591–609 (1984).
- Bergman, K. H. *Mon. Weath. Rev.* **112**, 1441–1456 (1984).
- Quiroz, R. S. *Mon. Weath. Rev.* **112**, 1894–1912 (1984).
- Wagner, A. J. *Mon. Weath. Rev.* **113**, 149–169 (1985).
- Ropelewski, C. F. *Mon. Weath. Rev.* **113**, 664–679 (1985).
- Keshavamurty, R. N. *J. Atmos. Sci.* **39**, 1241–1259 (1982).
- Blackmon, M. L., Geisler, J. E. & Pitcher, E. J. *J. Atmos. Sci.* **40**, 1410–1425 (1983).
- Shukla, J. & Wallace, J. M. *J. Atmos. Sci.* **40**, 1613–1630 (1983).
- Weare, B. C. *Q. Jl R. Met. Soc.* **103**, 467–478 (1977).
- Merle, J., Fieus, M. & Hisgard, P. *Deep Sea Res.* **26** (Suppl. II), 77–102 (1979).
- Weare, B. C., Navato, A. R. & Newell, R. E. *J. Phys. Oceanogr.* **6**, 671–678 (1976).
- Horel, J. D. *Mon. Weath. Rev.* **110**, 1863–1878 (1982).
- Reynolds, R. W. *NOAA Tech. Rep.* **31** (1982).
- Philander, S. G. H. *Nature* **322**, 236–238 (1986).
- Lamb, P. J., Pepler, R. A. & Hastenrath, S. *Nature* **322**, 238–243 (1986).
- Palmer, J. N. *Nature* **322**, 251–253 (1986).
- Hisard, P., Henin, C., Houghton, R., Piton, B. & Rual, R. *Nature* **322**, 243–245 (1986).
- Katz, E. J., Hisard, P., Verstraete, J. M. & Garzoli, G. L. *Nature* **322**, 245–247 (1986).
- Weisberg, R. W. & Colin, C. *Nature* **322**, 240–243 (1986).
- Kousky, V. E., Kagano, M. T. & Cavalcanti, I. F. A. *Tellus* **36A**, 490–504 (1984).
- Hastenrath, S. & Heller, L. *Q. Jl R. Met. Soc.* **103**, 77–92 (1977).
- Lamb, P. J. *Mon. Weath. Rev.* **106**, 482–491 (1978).

Influence of the Atlantic, Pacific and Indian Oceans on Sahel rainfall

T. N. Palmer

Meteorological Office, Bracknell, Berkshire RG12 2SZ, UK

Folland *et al.*¹ have reported that persistently dry and wet periods of several years in the Sahel have been accompanied by global-scale patterns of sea-surface temperature (SST) anomaly. They also demonstrated that the response of a general circulation model (GCM) of the atmosphere to an observed composite SST difference field between a number of such dry and wet periods showed substantial reduction in Sahel rainfall compared with values from a simulation with climatological SSTs. I examine here the same model's response to the individual components of the composite SST difference field in the Atlantic, Pacific and Indian Oceans. It is found that over the western Sahel, the Atlantic and Pacific fields have a comparable effect in reducing rainfall whereas the Indian Ocean field produces a slight enhancement. Results suggest that, over the eastern Sahel, the Indian Ocean has the dominant role in reducing rainfall.

Figure 1 shows the worldwide SST difference field between the five driest and five wettest years in the Sahel after 1949 and is discussed in more detail in ref. 1. Note that whilst this field

does represent differences between extreme years, values in the South Atlantic are not as large as anomalies (>2 K) observed during the exceptional warm event of 1984 (ref. 2). Similarly, values over the east Pacific are not as large as observed anomalies following the 1982/83 El Niño event³.

Results were obtained from several 180-day fixed July integrations of the Meteorological Office 11-layer GCM⁴ using the full difference field in Fig. 1, and its components in the three principal oceans. The integrations are identified as follows. The control integration has fixed climatological average July SSTs, and the anomaly integration (TOT) has the full field in Fig. 1 added to these values. Integrations ATL, PAC and IND have only those SSTs in Fig. 1 from the Atlantic, Pacific and Indian Oceans respectively, added to climatological values (see ref. 1 for a description of rainfall and low level moisture flux in the control integration).

Figure 2 shows 180-day mean-rainfall difference fields between the four anomaly integrations and the control. As discussed in ref. 1, in TOT (Fig. 2a) there is a coherent and substantial reduction in rainfall across the Atlantic and the Sahel, by up to 50% of the model's climatological mean. For integration ATL (Fig. 2b), the pattern of rainfall reduction over the Atlantic and the Western Sahel is similar to, but ~30% weaker than, that for integration TOT. Over the eastern Sahel, the reduction in rainfall is much less substantial in ATL.

For integration PAC, surprisingly at first sight, there is a similar pattern of rainfall reduction to that in TOT and ATL. Over the Atlantic, values are smaller than in ATL, and the rainfall enhancement to the south is absent. However, inland, over the Sahel, rainfall reduction seems to be larger than that in ATL.

Results from IND are different. Inland, over the western and central Sahel, an enhancement in rainfall of over 1 mm per day is apparent. Further east, over Sudan and northern Ethiopia, there is a decrease in rainfall by up to 50% of climatological values. Finally, over the western Indian Ocean itself, values are strongly enhanced. (Although not shown, rainfall over the eastern Indian Ocean is also increased.)

An assessment of whether these rainfall anomalies are significant was made by calculating a *t*-variant from the six pairs of non-overlapping 30-day mean values of the control and anomaly integrations. If the conditions of the *t*-test were satisfied, then anomalies with $|t| > 2$ would be accepted at about the 7% significance level, using a two-sided test. *t*-Values with magnitude > 2 shown stippled in Fig. 2. The reduction in rainfall across the Atlantic and into the extreme western Sahel is extremely-reproducible for experiment TOT whilst values of *t* are smaller in this region for ATL and, except for a small neighbourhood of the West African coast, $t < 2$ in PAC. In experiment IND, the increase in rainfall over the Indian Ocean is significant, as is the principal area of decrease over Sudan. In TOT, and more so in ATL, the increase in rainfall over north-east Brazil appears to be highly reproducible (although the northern summer is not the main rainy season in this area).

The addition of the SST anomaly in ATL reduces the meridional gradient of Atlantic SST just south of the inter-tropical convergence zone (ITCZ). As a result, the Atlantic Hadley cell south of the ITCZ is weakened and, as Fig. 3 shows, the low-level south-east Trade Winds are weakened. The flux of moisture into the ITCZ is reduced, producing less latent heating there, consistent with the weaker Hadley cell.

Whilst the main change in precipitation occurs over the Atlantic, Fig. 3a shows that reduction in low-level flow into the ITCZ also reduces the south-west monsoon flow into the western Sahel.

As mentioned above, the composite positive SST anomaly in Fig. 1 is considerably weaker, particularly in the Gulf of Guinea, than SST anomalies observed during the exceptional warming of 1984. Hence, it is very likely that a more marked weakening of the south-west monsoon flow into the western Sahel would

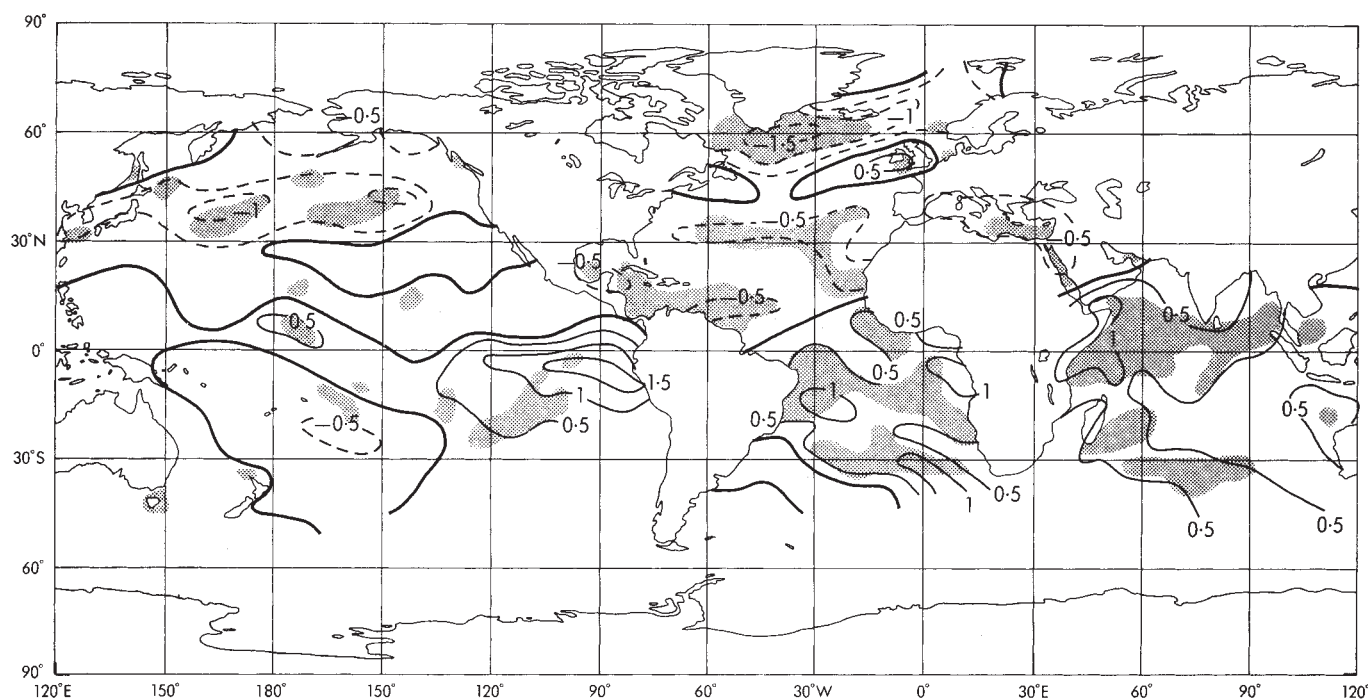


Fig. 1 SST for July to September. Average of 1972, 1973, 1982, 1983, 1984 (dry years) relative to average of 1950, 1952, 1953, 1954, 1958 (wet years). Contour interval, 0.5 °C. Shaded areas are different from zero at the 90% level of significance according to a *t*-test.

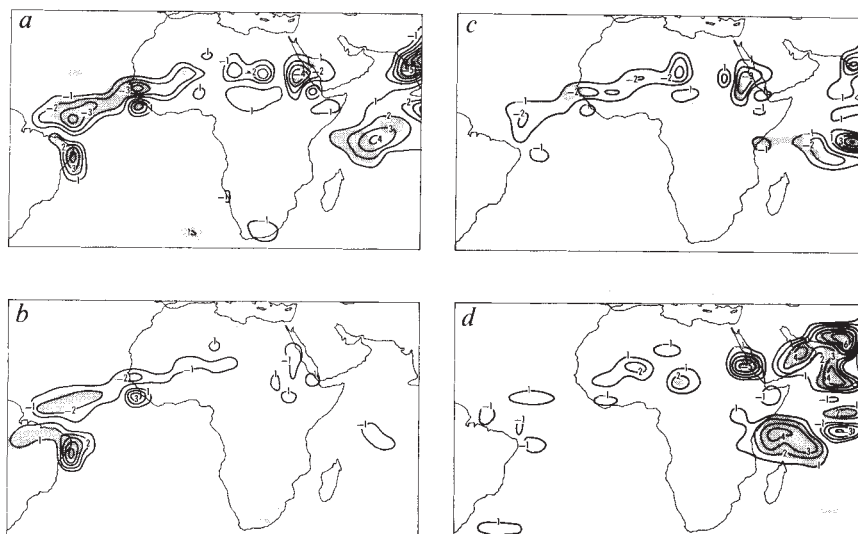


Fig. 2 Rainfall anomaly (relative to control integration) for experiment with: *a*, Full SST anomaly in Fig. 1 (TOT); *b*, Atlantic portion of Fig. 1 only (ATL); *c*, Pacific portion of Fig. 1 only (PAC); *d*, Indian Ocean portion of Fig. 1 only (IND). Contour interval, 1 mm per day. Stippling indicates areas where the magnitude of the *t*-variate exceeds 2.

have resulted if the 1984 SST anomalies had been used in place of those in Fig. 1. We hope to test this speculation in a future integration.

Whilst the SST anomaly field in Fig. 1 has negative values over 1 K in the north Pacific, the principal influence of SST anomalies in PAC on rainfall is over the tropical Pacific. With water at 28 °C or warmer extending across most of the tropical Pacific north of the Equator in the climatological normal, the anomaly near the dateline in Fig. 1 produces a substantial increase up to 10 mm per day over the central and west Pacific, whilst anomalies west of Central America (which weaken the meridional SST gradient there) are associated with a reduction of the maximum of rainfall west of Mexico by up to 10 mm per day. Previous GCM studies with El Niño SST anomalies⁵ have

shown that enhanced latent heat release in the central equatorial Pacific excites an atmospheric Kelvin wave which can propagate across the Atlantic and Africa without complete dissipation. As a result the equatorial upper troposphere is warmed, geopotential height is raised and anomalous equatorial westerlies are produced, in agreement with results from thermally forced linear equatorial β -plane models⁶. Figure 3*b* shows the 250-mbar streamfunction anomaly from PAC. The north-to-south gradient indicates that the anomalous non-divergent wind in the tropics at 250 mbar indeed has a westerly component, strongest over the east Pacific, but extending over the Atlantic and Africa. The wind anomalies have a baroclinic vertical structure near the ITCZ over the Pacific, Atlantic and Sahel, with an easterly component at low levels.

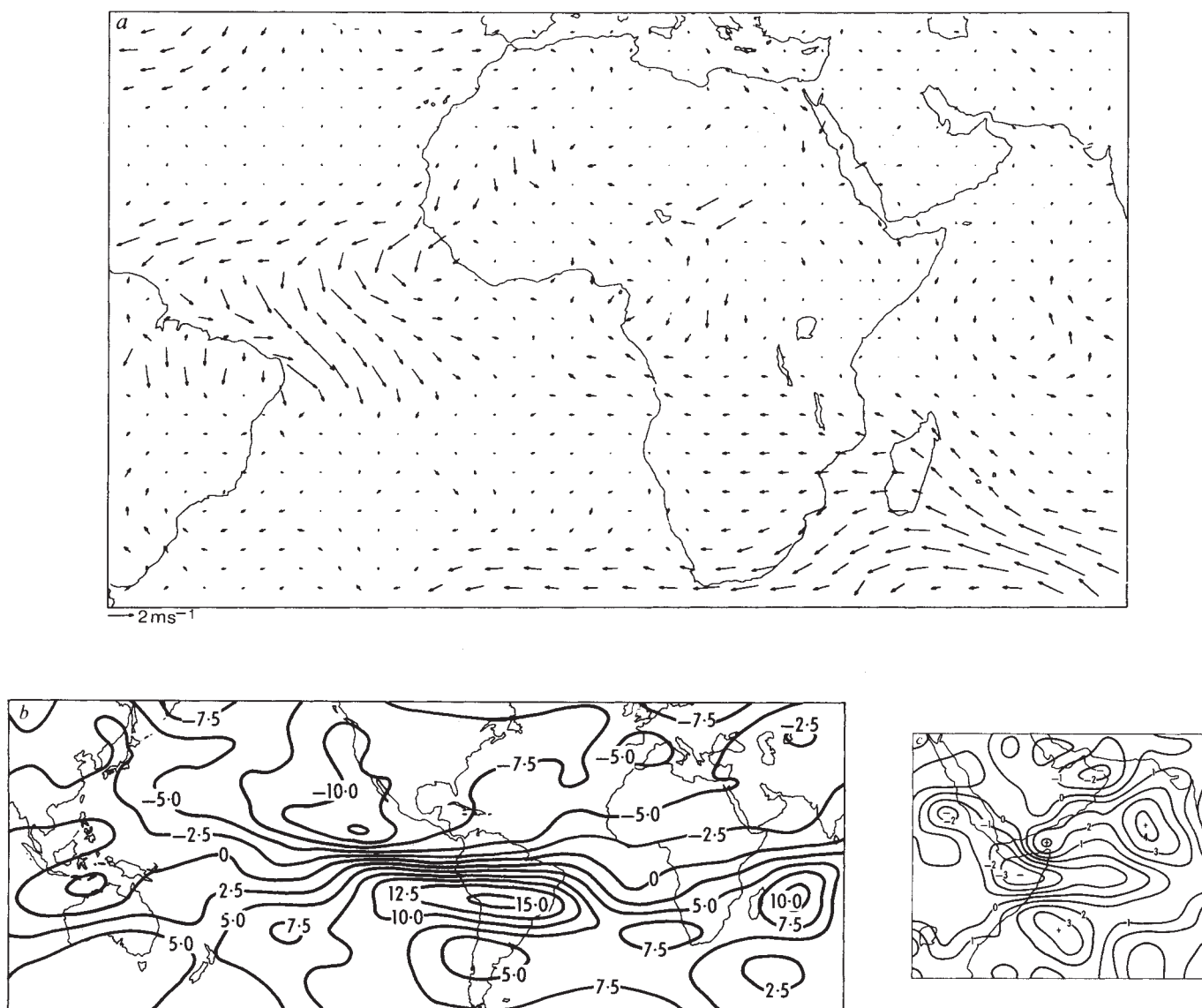


Fig. 3 a, 950-mbar wind anomaly for experiment ATL; b, 250-mbar streamfunction anomaly ($\times 10^6 \text{ m}^2 \text{ s}^{-1}$) for PAC; c, 250-mbar divergence anomaly ($\times 10^{-6} \text{ s}^{-1}$) for IND.

This again reduces the south-west monsoon flow into the western Sahel, and thereby the supply of moisture for precipitation. The fact that there is no significant increase in rainfall over the Atlantic south of the ITCZ in PAC is consistent with the divergent circulation anomalies being mainly east/west oriented ('Walker cell') rather than north/south ('Hadley cell') as in ATL.

Results from IND are more easily understood. In Fig. 2d, there was substantial enhancement of rainfall over the Indian Ocean; Figure 3c shows anomalous divergence at 250 mbar associated with this enhancement. The correspondence between divergence and rainfall arises from the local thermodynamic balance in the tropics⁷ between adiabatic cooling by vertical motion and latent heat release. The main area of descent required by mass continuity appears to be located over Ethiopia and Somalia, consistent with the rainfall decrease there. The anomalous low-level wind in this region (not shown) flows from land to ocean, similar to Fig. 3a for ATL.

The tropical Atlantic therefore has a role in modulating rainfall over the Sahel, and was probably paramount during the exceptional 1984 season. On the other hand, effects from either the Pacific or Indian Ocean cannot be ignored. Hence, a com-

plete understanding of variability of rainfall in the Sahel involves consideration of the entire tropical circulation. As such, these model results are in agreement with the observational correlations between worldwide SST and Sahel rainfall, discussed by Folland *et al.*¹

However, results from the *t*-test also show that over most of the Sahel, rainfall anomalies were quite variable. On the other hand, a stronger and more significant response to SST anomalies in this area might be expected if coupling between rainfall, ground reflectivity and soil moisture capacity⁴ were built into the model.

C. K. Folland and D. E. Parker supplied the SST data shown in Fig. 1. I also thank them for many useful discussions.

Received 4 November 1985; accepted 2 April 1986.

1. Folland, C. K., Palmer, T. N. & Parker, D. E. *Nature* **320**, 602-607 (1986).
2. Philander, S. G. H. *Nature* **322**, 236-238 (1986).
3. Gill, A. E. & Rasmusson, E. M. *Nature* **306**, 229-234 (1983).
4. Cunningham, W. M. & Rowntree, P. R. *Q. Jl R. met. Soc.* (in the press).
5. Palmer, T. N. in *Coupled Ocean Atmosphere Models* (Elsevier, Amsterdam, 1985).
6. Gill, A. E. *Q. Jl R. met. Soc.* **444**-462 (1980).
7. Webster, P. J. *J. atmos. Sci.* **38**, 554-571 (1981).

River dynamics and the diversity of Amazon lowland forest

Jukka Salo*, Risto Kalliola*, Ilmari Häkkinen*,
Yrjö Mäkinen*, Pekka Niemelä*, Maarit Puhakka*
& Phyllis D. Coley†

* Department of Biology, University of Turku,
SF-20500 Turku, Finland

† Department of Biology, University of Utah, Salt Lake City,
Utah 84112, USA

We suggest here that large-scale natural forest disturbance and primary succession in the lowland rainforests of the Peruvian Amazon is caused by lateral erosion and channel changes of meandering rivers. Our results indicate that in the upper Amazon region, primary succession on newly deposited riverine soils is a major mode of forest regeneration. Landsat imagery analyses show that 26.6% of the modern lowland forest has characteristics of recent erosional and depositional activity; 12.0% of the Peruvian lowland forest is in successional stages along rivers. This successional development is used to classify the western Amazon rainforests according to their geomorphological erosion-deposition pattern. We also propose that by causing high site turnover, disturbance and variation in forest structure, the river dynamics may be a major factor creating and maintaining the high between-habitat (β -type) species diversity characterizing the upper Amazon¹.

The degree of forest disturbance caused by river-channel erosion was studied in $>0.5 \times 10^6$ km² of the Peruvian Amazon by using analyses of Landsat MSS (multi-spectral scanner) images (visual analysis). The disturbance is caused by the lateral erosion and channel changes of the meandering and braided white-water rivers (rich in fine suspended inorganic solids; see refs 2-7) of the western periphery of the Amazon region⁸⁻¹². Patterns of primary succession on newly deposited riverine soils were examined in more detail along the Rio Madre de Dios and Rio de los Amigos in the Madre de Dios region of southeastern Peru in 1982. The work was continued at the Cocha Cashu biological station at Manú National Park in 1983-86.

Twenty-two Landsat images of the Peruvian headwater region were surveyed for indications of former riverine erosion (Fig. 1, Table 1). In the analyses, we defined three erosional-depositional formations: (1) present floodplain formation, (2) previous floodplain formation, unquestionable floodplain origin; and (3) forests on denuded soils, which includes all the forests beyond the present or previous floodplains (these are probably also of former floodplain/lacustrine origin^{13,14}, derived from the depositional processes described here).

(1) Forests on the present floodplains are the areas most recently subjected to lateral erosion and channel migration. They were detected on the images by delimiting the areas covered by the present meander plains, which were defined as the areas containing the present meanders and the most recent channel cutoffs and oxbow lakes. This area, which has recently undergone several cycles of complete lateral erosion, may be further divided into two distinct types: sequential successional forest and mosaic forest (Fig. 2). Active deposition occurs on the inside of the meander bends, by point bar accretion, and with time meander migration has produced a series of meander scrolls; these form a sequence of ridges running parallel to the river course, on which a sequential primary forest succession takes place. The sequential succession proceeds undisturbed until the migrating river channel cuts the meander from a different angle.

According to the repetitive nature of river dynamics, the migration of the river channel course creates a mosaic of successional forests within the present meander plain. The mosaic forest is composed of patches of differentially aged sequential

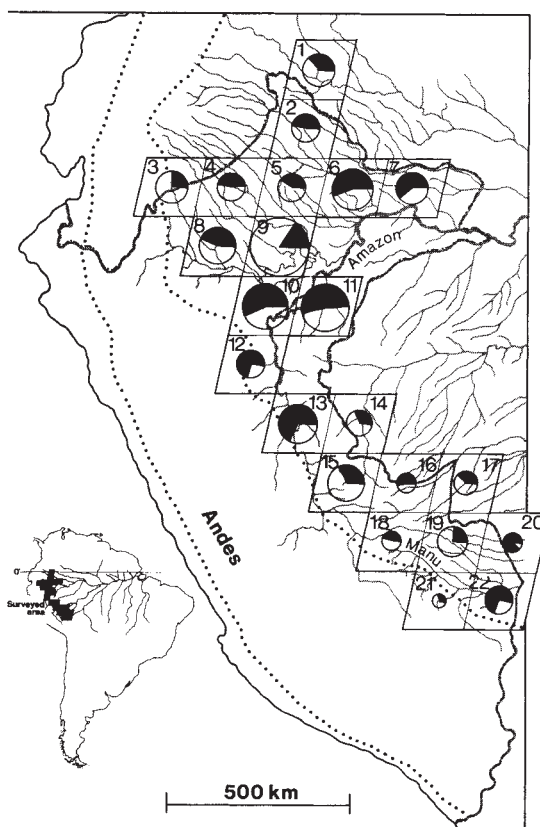


Fig. 1 The distribution and areas of riverine disturbance in the Peruvian Amazon basin. Each of the 22 squares represents one Landsat image (182.5×182.5 km). The sizes of the circles give the area (km²) of total floodplain (see text and Table 1 for details), representing the disturbed area covered by sequential floodplain generations. The black sectors represent the area of the forests on the present floodplains (sequential successional and mosaic forests). The white sectors show the area of the forests on the previous floodplains, indicating transitional forest. Inset, orientation map.

successional forests and patches of forests originating from a succession on the sites of former oxbow lakes. The annual floods further modify the mosaic pattern. The total area covered by the present floodplains is 12.0% of the area analysed.

(2) The area covered by former floodplain formations was defined by identifying the areas of flat surface topography without the convex-concave morphology¹⁵ (indicating surface erosion; for details see refs 16-19) characteristic of the upper Amazon forests on the oldest alluvium of diverse nature. The surface processes acting on Amazon abandoned floodplains change from depositional to erosional with increasing age²⁰. The area characterized by lack of the convex-concave topography is also scattered with isolated oxbow lakes, thus confirming the fluvial/floodplain origin of this formation. We believe that the successional status of forests growing on the soils of previous floodplains is transitional, as the mode of forest regeneration takes place increasingly through secondary succession, promoted by light gaps.

Previous floodplain formations cover 14.6% of the area analysed. The geomorphology of the old floodplains remains highly heterogeneous and is composed of a mosaic of abandoned river channels, oxbow lakes and sedimentary beds of different ages. This heterogeneity gives rise to densely packed plant communities of different ages and origin. In contrast to forests under the influence of present floodplains, older areas of floodplain are scattered throughout the analysed area, indicating major channel changes in the recent history of the western Amazon river system.

(3) The area beyond the present and previous floodplain

Table 1 Peruvian Amazon basin areas subjected to riverine disturbance

No. in Fig. 1	Landsat scene			Analysed area*	Forests on present floodplains (sequential and mosaic forest)		Forests on previous floodplains (transitional forest)		Total floodplain	
	Path	Row	Date		Area (km ²)	(%)	Area (km ²)	(%)	Area (km ²)	% Of total
1	007	60	2 October 1972	29,900	2,050	6.9	3,650	12.2	5,700	19.1
2	007	61	2 October 1972	27,000	2,050	7.6	2,150	8.0	4,200	15.6
3	009	62	29 September 1973	15,700	1,350	8.6	4,650	29.6	6,000	38.2
4	008	62	24 July 1977	21,400	1,900	8.9	2,000	9.3	3,900	18.2
5	007	62	8 December 1973	24,400	1,900	7.8	2,700	11.1	4,600	18.9
6	006	62	7 December 1973	33,300	5,150	15.5	4,050	12.1	9,200	27.6
7	005	62	3 February 1973	26,300	4,000	15.2	2,300	8.7	6,300	23.9
8	008	63	10 September 1973	29,700	3,250	10.9	4,050	13.7	7,300	24.6
9	007	63	2 October 1972	24,000	2,950	12.3	15,950	66.5	18,900	78.8
10	007	64	8 December 1973	25,600	7,050	27.5	4,950	19.4	12,000	46.9
11	006	64	7 December 1973	30,300	9,800	32.3	8,000	26.4	17,800	58.7
12	007	65	2 October 1972	13,600	2,800	20.6	1,200	8.8	4,000	29.4
13	006	66	24 August 1981	20,600	6,600	32.0	3,200	15.5	9,800	47.6
14	005	66	11 July 1979	26,600	950	3.6	1,850	6.9	2,800	10.5
15	005	67	27 April 1976	16,600	2,150	12.9	4,450	26.8	6,600	39.7
16	004	67	5 September 1975	32,200	950	2.9	750	2.4	1,700	5.3
17	003	67	28 October 1975	26,000	1,050	4.0	1,650	6.3	2,700	10.3
18	004	68	7 July 1976	17,100	700	4.1	800	4.7	1,500	8.8
19	003	68	22 August 1976	25,400	1,250	4.9	4,350	17.1	5,600	22.0
20	002	68	29 July 1975	26,300	1,600	6.1	100	0.4	1,700	6.5
21	003	69	3 December 1975	4,000	300	7.5	900	22.5	1,200	30.0
22	002	69	29 July 1975	19,800	2,300	11.6	1,400	7.1	3,700	18.7
Total				515,800	62,100	12.0	75,100	14.6	137,200	26.6

Forests on the present floodplains are undergoing primary and early succession on newly deposited riverine soils. The present floodplain formation is composed of the sequential successional and mosaic forests. Forests on the previous floodplains represent the area covered by a series of older floodplain formations (transitional forest). The analyses are based on visual analysis of 1:10⁶ Landsat images MSS 7, except scenes 007-63 and 007-65, which are MSS 6.

* In the analysis, Andean rifts and areas shaded by clouds or other turbulence are excluded. The total area of each image is 33,300 km².

formations has the convex-concave surface erosion pattern. The origin and geomorphology of these areas is unknown, but the erosion pattern is found on both floodplain and lacustrine (Belterra clay type) deposits. As the processes operating on the floodplains change towards erosional, the topography becomes similar to that of the areas of convex-concave relief^{15,20} (see Fig. 2).

In the western Amazon, the structure of the denuded topmost sedimentary layer has been only locally surveyed^{8-12,21-24}. However, there is good reason to believe that after the glacial periods of the Quaternary, the Andean tills, re-worked by rivers, formed a belt of highly heterogeneous sedimentary beds in front of the foothills of the Eastern Cordillera²⁴. In Peru and Bolivia, the uppermost sedimentary bed, called the Iñapari Formation, overlies the Tertiary red beds²⁵⁻²⁸, especially at sites of high river erosion banks ('cerros'). Recent radiocarbon dates of four samples collected from three 35-m eroding channel banks along the Rio Acre, Bolivia, show that these sediments were deposited between 5,575 ± 105 and 10,085 ± 150 yr BP (ref. 28). These data indicate the youth of the Iñapari Formation, although the stratigraphy of the samples may merely reflect a local sedimentation pattern within the Rio Acre region.

To evaluate the rate of forest regeneration within the areas of sequential successional and mosaic forests, an analysis of a 70-km section of the Rio Manú from the Cocha Cashu station (11°53' S, 71°24' W; 350 m above sea level) to Boca Manú (12°15' S, 70°55' W, 320 m a.s.l.) was carried out. We compared 1962-63 aerial photographs with a 1976 Landsat image (276296-135752-7). The mean lateral erosion rate of the meander bends during the 13-yr period was 12 m yr⁻¹. During this period, the total area of newly created depositional land subject to primary succession was 12.0 km², representing 3.7% of the present floodplain formation area (325 km²).

The typical pattern of riparian primary succession (sequential successional forest) is summarized in Fig. 3, where characteristics of the successional meander forest which is common along most of the western Amazon lowland rivers²⁹ are emphasized. Extreme zonation of vegetation occurs during the early phases of succession (up to the *Ficus-Cedrela* zone³⁰). This zonation is initially created by the instability of the youngest point bar deposits ('beaches') on the inner curve of the meander (see Fig. 3 for details). *Tessaria integrifolia* (Asteraceae) is the first pioneer tree to appear, because its seedlings can tolerate the arid and unshaded conditions of this environment. Invasion by other successional trees is possible only after modification of the microclimate by *Tessaria*. At this stage, competition between species may further influence the zonation pattern.

The traditional view of the Amazon rainforests has emphasized stability, with the dominant mode of forest regeneration occurring in light gaps created by fallen trees. Estimated tree-fall frequencies from several lowland tropical forests indicate that 3-5% of the forest is in the first 5 years of regeneration in light gaps³¹⁻³³. Our results show that the lateral erosion and channel changes of rivers repeatedly disturb all types of forest in the western Amazon lowlands, and thus create a mosaic of forests with higher age heterogeneity than could have resulted from regeneration in light gaps alone. Also, the nature of forest regeneration is fundamentally different in these two processes. The erosional disturbance is much more severe than disturbances caused by tree-falls. Regeneration on newly exposed fluvial deposits ('beaches') must start with the earliest phases of primary succession. Regeneration in tree-falls is due to growth of established saplings or of a well-developed seed bank. Finally, areas of riparian succession occur in long, continuous and dynamic strips along recent and abandoned river channels. This spatial

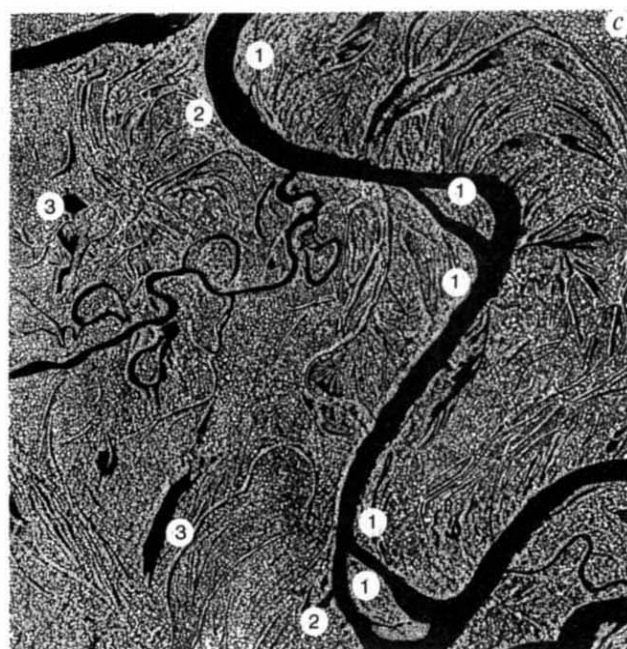
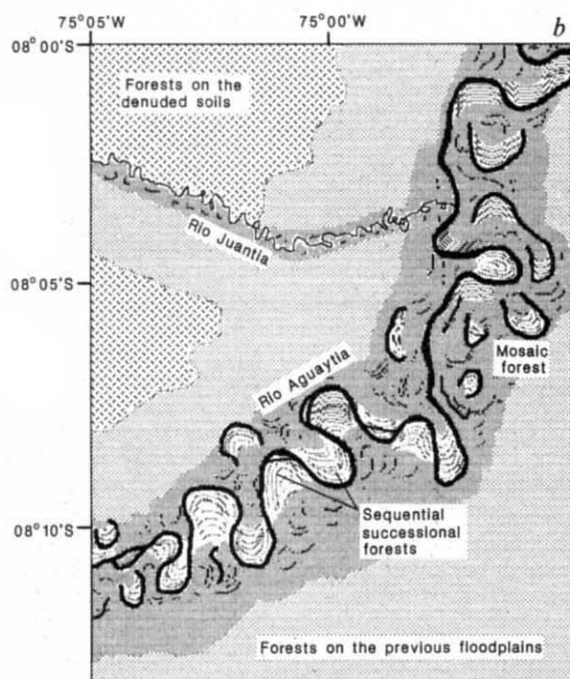
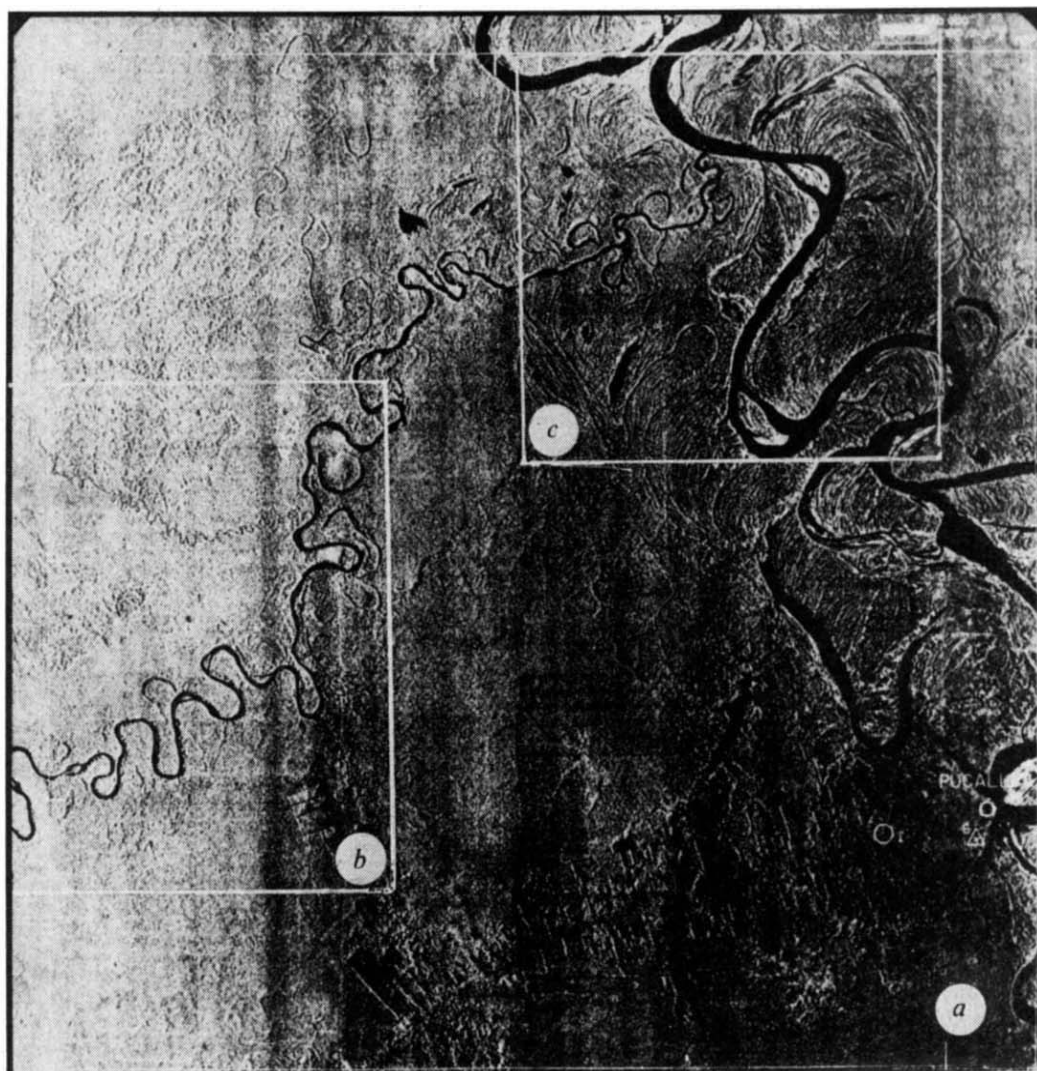


Fig. 2 (Left) *a*, Side-looking airborne radar (SLAR) image of the floodplain mosaic formed by the white-water river Ucayali at Pucallpa, Perú (8°15' S, 74°40' W). *b*, Simplified map of the forest disturbance site in *a*. The sequential successional forest is created on both sides of the meandering rivers Aguaytía and Juntía. As the meandering proceeds, the sequential successional forest is cut from different angles and a mosaic forest structure is created. During the annual flood cycle, sedimentation modifies the sequential meander scroll topography towards the flat topography of the former floodplain area. The previous floodplains are generally at a topographically higher level than the present floodplains. The denuded forest areas (upper left corner) representing the surface erosion relief pattern (convex-concave morphology) are being laterally eroded by the Rio Juntía. The transitional forest (lower right corner) is developing towards the denuded pattern. *c*, Detailed map of forest ground development. 1, Sites of intense primary succession at meander points leading to sequential successional forest. 2, Mosaic and transitional forest subject to lateral erosion at the outer curve of the meanders. 3, Isolated oxbow lakes. Based on ONERN SLAR image SC-18 3.A.

distribution differs markedly from the scattered distribution of light gaps.

We also propose that forest disturbance due to modern and past river dynamics is partially responsible for the high biological diversity in the upper Amazon basin³⁴⁻³⁶, for several reasons. First, many distinct habitats are generated by erosional and depositional activity of rivers. As the forest succession and river bank erosion proceed, a mosaic of forests of different ages and on different soils is created (Fig. 2). A good example of this is the species richness of the Tambopata and Manú reserves in southeastern Peru^{30,36}, both of which are located at sites which have been, and still are, modified by relatively small meandering rivers (Rio Tambopata and Rio Manú, respectively) causing closely packed, small-scale habitat mosaics. Second, the habitats are highly stable in species composition and are relatively short-lived, so there is insufficient time for competitive exclusion³⁷. Finally, the floodplains may differ profoundly in water and soil chemistry, mode of alluvial sedimentation, and case-historical bio-geographical events.

The river dynamics may also represent a mechanism leading to a pattern of equal allopatric speciation, as has been proposed to operate in the context of the Pleistocene refuges³⁸. The Pleistocene refuge theory states that the basis for the prevailing high species richness of the Amazon lowlands was laid down during the arid phases of the Pleistocene³⁹. During this time, the continuous forest was repeatedly fragmented into savannah-bordered isolates. Owing to the dynamic nature of the topmost sedimentary layer in the western Amazon, both floodplains separated by old denuded-soil forests and the denuded-soil

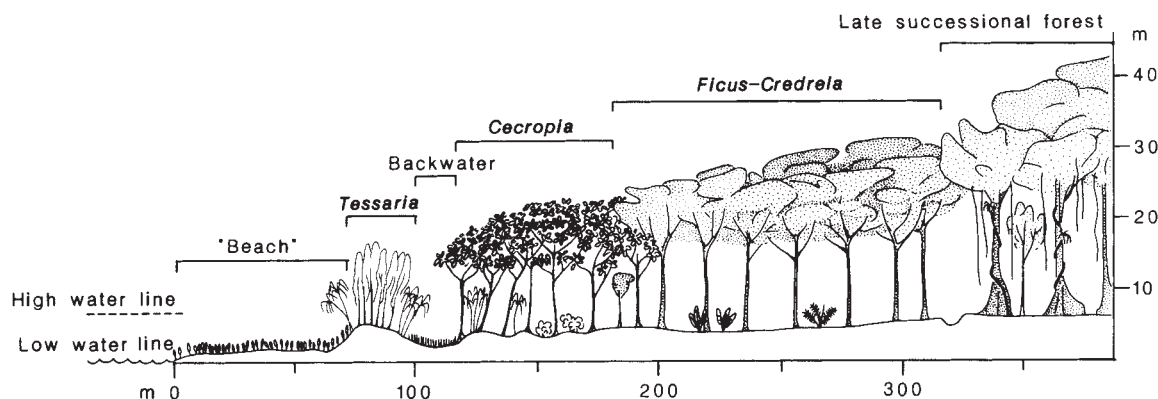


Fig. 3 Riparian succession in the sequential successional forest next to the Rio Manú at the Cocha Cashu biological station (11°53' S, 71°24' W). The scheme is a simplified transect of the Cocha Cashu meander loop from the point (left) towards the neck. The 'beach' is created by the deposition of river sediment on the point bar of the meander bank. The dominating species (~99%) is *Tessaria integrifolia* (Asteraceae). The persistence of *Tessaria* seedlings depends on the stability of the 'beach'; successful colonization reaching maturity is expected to take place only once in a 4-yr period. Other primary successional species include: *Baccharis salicifolia* (Asteraceae), *Cassia reticulata* (Leguminosae), *Cyperus* spp. (Cyperaceae), *Ipomoea* spp. (Convolvulaceae), *Ochroma lagopus* (Bombacaceae) and *Panicum* spp. (Poaceae). The *Tessaria*-*Gynerium* zone is situated on the youngest point bar, the topography of which is accentuated by the deposition of sandy river sediment during the annual flooding of the river, predominantly during December–April. In the middle part of the zone, stands of reproductive *Tessaria* of equal age predominate. Both sides of the zone are dominated by *Gynerium sagittatum* (Poaceae). Gradually, *Tessaria* is replaced by *Gynerium*. The zone harbours seedlings of *Cecropia* cf. *tessmannii* (Moraceae), *Cedrela odorata* (Meliaceae) and *Ficus insipida* (Moraceae). The 10–30-m-wide backwater depression marks the previous position of the river channel, before accretion of the youngest point bar. During the high-water period the depression is often used as a secondary channel, or chute, as a result of turbulence in the flood waters. The backwater depressions run parallel to the river and their number in meander loops varies according to the local turbulence patterns. In the middle of the backwater depression runs a belt of *Hymenachne donacifolia* (Poaceae). The edges of backwater depressions are invaded by young *Cecropia* trees. As the backwater belt becomes subject to lesser annual flooding, the *Cecropia* forms a closed canopy over the belt. Before this, the stagnant water prevents the invasion of other trees or *Gynerium*. The *Cecropia* zone is located on the older meander scrolls with intermediate former backwater depressions. The dominating soil type is still fluvial sand. The forest is dominated by *Cecropia tessmannii*, forming an open canopy at 14–18 m. The gaps in the canopy promote patches of declining *Gynerium*. The gaps are filled by *Casearia* sp. (Flacourtiaceae), *Cedrela odorata*, *Erythrina* spp. (Leguminosae), *Ficus insipida*, *Guarea* sp. (Meliaceae) and *Sapium* sp. (Euphorbiaceae). The understory includes members of the families Euphorbiaceae, Melastomataceae, Piperaceae and Acanthaceae. The topography in the *Ficus*-*Cedrela* zone is more gentle than in the previous stages due to the sedimentation of alluvial silt which is deposited annually during the flooding period, forming a layer above the riverine sands of the meander scrolls and chutes. The zonation characterizing the vegetation in the previous phases is diminishing. The closed forest canopy is formed mainly by fruiting *Ficus insipida* and *Cedrela odorata*. The understory is dominated by thick growth of *Heliconia* spp. (Musaceae). The *Ficus*-*Cedrela* forest is characterized as the first zone containing palms. The topography of the late successional forest (mosaic forest) is generally flat, with minor depressions and gentle hills. The forest structure is further diversified by autogenic succession promoted by light gaps.

forests themselves may act as isolates within which biological differentiation can occur. There is a growing body of evidence that the number of species confined only to floodplain or denuded-soil forests is exceptionally high in the Amazon⁴⁰⁻⁴³. If the rivers repeatedly form new major floodplains and abandon old ones, subsequent secondary contact of formerly separated river systems, floodplains and denuded-soil forests may lead to species dynamics similar to the refuge pattern.

The ecological consequences of river dynamics and riparian primary succession also have profound implications for the conservation policies of the Amazon. As a major part of the forest has gone through riparian succession, the first phases of succession may be essential in determining the structure of the later successional forest. Conservation programmes based solely on the Pleistocene refuge theory may not preserve the present dynamic mechanism underlying the maintenance of high species diversity.

We thank F. S. Chapin, D. Davidson, I. Hanski, E. Haukioja, S. Hinneri, D. Janzen, O. Järvinen, K. J. Laine, O. V. Lindqvist, S. Neuvonen, D. Pearson, M. Räsänen, J. Tuomi, K. Viherkoski, P. Voss and T. Vuorisalo for constructive comments. J. Terborgh, L. Emmons and other workers at the Cocha Cashu biological station, Peru, provided invaluable help during the field work. T. Laurikainen and M. Mustonen drew the figures. ONERN, Lima, provided the SLAR and aerial photos of the monitored sites. The research was sponsored by the Academy of Finland under agreement 29/103 and by the Turku University Foundation.

Received 26 November 1985; accepted 23 April 1986.

Normal maturation involves systematic changes in binocular visual connections in *Xenopus laevis*

S. Grant* & M. J. Keating†

National Institute for Medical Research, Mill Hill, London NW7 1AA, UK

Systematic changes in neuronal connections have been observed during the development of many vertebrate neuronal systems. These changes have usually involved a refinement from an initial exuberance of connections¹⁻⁴ or a response to some experimental perturbation⁵⁻⁸. Here we report on a system of neuronal connections, which, during a protracted developmental period, undergo ordered changes in response to normally occurring changes in functional requirements. In the frog *Xenopus laevis*, interocular alignment changes markedly during late larval and post-metamorphic life, producing a progressive enlargement of the binocular portion of the visual field^{9,10}. An intertectal system links the two mid-brain optic tecta and is concerned with the neural representation of binocular visual space. In the adult animal, connections in this system link corresponding points (points receiving information from one locus of binocular visual space) on the two tecta. Changes in eye position with development, however, change the set of corresponding points. Therefore, if the intertectal connections link corresponding tectal points throughout development, they must undergo an ordered change with time. We present electrophysiological evidence that the intertectal connections do, indeed, undergo such changes in response to changes in eye alignment, and that the changes are major.

In the late larval *Xenopus* the two eyes face laterally and there is no fronto-superior binocular visual field. As metamorphic climax approaches, changes in skull shape produce frontal and dorsal migration of the two eyes. This movement persists into post-metamorphic and adult life. Optical methods were used to determine the relative alignment of the two optic axes and the extent of the two monocular visual fields^{9,10}. These measurements indicated that the interocular angle changes from 160° at stage 60 (ref. 11), the onset of metamorphic climax, to 50° in adult life. The monocular visual field remains relatively constant between 195° and 200° throughout this period. The binocular visual field (the area of overlap of the two monocular visual fields) thus increases from ~35° at stage 60 to 45° at stage 62, 100° at stage 66 (the completion of metamorphic climax) and ~150° in adult life. This growth of the binocular visual field with age is plotted, on right eye-centred coordinates, in Fig. 1.

In the adult each optic tectum receives two projections of binocular visual space, one through each eye. The contralateral visuotectal projection is the product of the direct retinal fibre input from an eye to its contralateral optic tectum. The system of intertectal connections conveys visual information from the contralateral to the ipsilateral tectum and thus underlies the ipsilateral visuotectal projection¹²⁻¹⁸. Figure 2 shows the four visuotectal projections in a normal adult. The adult ipsilateral visuotectal projection displays several characteristics: it arises from the full extent of the binocular visual field, it is topographically ordered, and is in register with the contralateral visuotectal projection to that optic tectum. Figure 2 legend outlines the analysis by which the pattern of functional intertectal connections may be determined from the visuotectal projections.

To determine whether the pattern of intertectal connections changes during normal development, visuotectal projections were mapped electrophysiologically in 52 animals from stage

* Present address: Department of Anatomy, Medical College of Pennsylvania, 3200 Henry Avenue, Philadelphia, Pennsylvania 19129, USA.

† To whom correspondence should be addressed.

- Gentry, A. H. *Pl. Syst. Evolut.* **137**, 95-105 (1981).
- Sioli, H. *Forsch. Fortschr.* **26**, 274-280 (1950).
- Sioli, H. *Geol. Rdsh* **45**, 608-633 (1957).
- Sioli, H. *Amazoniana* **1**, 74-83 (1965).
- Fittkau, E. J. *Münster. Forsch. Geol. Paläont.* **20/21**, 35-50 (1971).
- Meade, R. H., Dunne, T., Richey, J. E., de M. Santos, U. & Salati, E. *Science* **228**, 488-490 (1985).
- Junk, W. J. & Furch, K. in *Amazonia* (eds Prance, G. & Lovejoy, T. E.), 3-17 (Pergamon, Oxford, 1985).
- Liu, K.-b. & Colinvaux, P. A. *Nature* **318**, 556-557 (1985).
- Servant, M. & Villarroel, R. *C. r. heb. Séanc. Acad. Sci., Paris* **288**, 665-668 (1979).
- Servant, M., Fontes, J.-C., Argollo, J. & Saliège, J.-F. *C. r. heb. Séanc. Acad. Sci., Paris* **292**, 1209-1212 (1981).
- Servant, M., Fontes, J.-C., Rieu, M. & Saliège, J.-F. *C. r. heb. Séanc. Acad. Sci., Paris* **292**, 1295-1297 (1981).
- Soubies, F. *Cah. ORSTOM, sér. Géol.* **XI**, 133-148 (1980).
- Putzer, H. in *The Amazon, Limnology and Landscape Ecology of a Mighty Tropical River and its Basin* (ed. Sioli, H.) 15-46 (Junk, Dordrecht, 1984).
- Irion, G. in *The Amazon* (ed. Sioli, H.) 201-214 (Junk, Dordrecht, 1984).
- Eden, M. J., McGregor, D. F. M. & Morelo, J. A. Z. *Geomorph.* **26**, 343-363 (1982).
- Sombroek, W. G. *Amazon Soils* (Centre for Agricultural Publications and Documentation, Wageningen, 1966).
- Tricart, J. *Annls Géogr.* **84**, 1-54 (1977).
- Mousinho de Meis, M. R. *Revta bras. Geogr.* **30**, 3-20 (1968).
- Colinvaux, P. A., Miller, M. C., Liu, K.-b., Steinitz-Kannan, M. & Frost, I. *Nature* **313**, 42-45 (1984).
- Klammer, G. in *The Amazon* (ed. Sioli, H.) 47-83 (Junk, Dordrecht, 1984).
- Rüegg, W. & Rosenzweig, A. *Soc. Geol. Perú Vol. Jubilar* **113**, 1-24 (1949).
- Gibbs, R. J. *Bull. geol. Soc. Am.* **78**, 1203-1232 (1967).
- Irion, G. in *Neotropische Ökosysteme* (ed. Müller, P.) 163-167 (Junk, The Hague, 1976).
- Putzer, H. *Mitt. Geol. Inst. TU Hannover* **8**, 87-96 (1968).
- ONERN *Inventario, Evaluación e Integración de los Recursos Naturales de la Zona de los Ríos Inambari y Madre de Dios* (Lima, 1972).
- ONERN *Inventario, Evaluación e Integración de los Recursos Naturales de la Zona Iberia-Iñapari* (Lima, 1977).
- Paula Couto, C. de *Iheringia* **5**, 3-17 (1978).
- Campbell, K. E. & Frailey, D. *Quat. Res.* **21**, 369-375 (1984).
- Ule, E. *Bot. Jb.* **40**, 114-172, 398-443 (1908).
- Terborgh, J. *Five New World Primates* (Princeton University Press, 1983).
- Hartshorn, G. S. *Biotropica Suppl.* **12**, 23-30 (1980).
- Uhl, C. & Murphy, P. G. *Trop. Ecol.* **22**, 219-237 (1982).
- Dourojeanni, M. J. R. & Ponce, C. F. in *Los Parques Nacionales del Perú* (eds Dourojeanni, M. J. R. & Ponce, C. F.) 176-201 (Infaco, Madrid, 1978).
- Encarnación, F. *Proyecto PNUD/FAO/PER/81/002 7*, 1983.
- Lao, R. *Rev. Forest. Perú* **3**, 1-61 (1969).
- Brown, K. S. *News Lepid. Soc.* **1984**, 45-47 (1984).
- Connel, J. H. *Science* **204**, 1344-1345 (1978).
- Haffer, J. *Science* **165**, 131-137 (1969).
- Prance, G. T. in *Amazonia* (eds Prance, G. T. & Lovejoy, T. E.) 146-165 (Pergamon, Oxford, 1985).
- Kinzey, W. G. & Gentry, A. H. in *Primate Ecology: Problem-Oriented Field Studies* (ed. Sussman, R. W.) 89-100 (Wiley, New York, 1979).
- Remsen, J. V. & Parker, T. A. *Biotropica* **15**, 223-231 (1983).
- Erwin, T. & Adis, J. in *Biological Diversification in Tropics* (ed. Prance, G. T.) 358-371 (Columbia University Press, New York, 1982).
- Pires, J. M. in *The Amazon* (ed. Sioli, H.) 581-602 (Junk, Dordrecht, 1984).

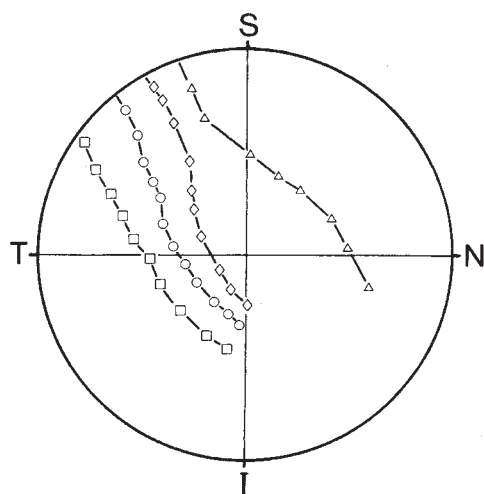


Fig. 1 Developmental changes in the extent of the binocular visual field in *Xenopus*. For these measurements, animals were placed at the centre of an Aimark projection perimeter (radius 33 cm) with the optic axis of the right eye aligned with the fixation point of the perimeter. The larger circle represents the visual field of the right eye, subtending some 200° of solid angle about the origin of this polar coordinate representation of visual space. The animal is thus positioned such that the right eye is looking out from the surface of the page. N, T, S and I are the nasal, temporal, superior and inferior poles of the visual field of the right eye. Reversible ophthalmoscopic methods used to determine the boundaries of the two monocular visual fields were those described by Grobstein and Comer⁹. The symbols represent the optically-determined naso-superior boundary of the visual field of the left eye at different developmental stages. The results are presented as mean observations from animals at stage 62 (Δ , $n=2$), stage 66 ($\diamond$, $n=3$), 3 months after metamorphosis ($\circ$, $n=3$) and adulthood ($\square$, $n=4$). Neither the range nor standard error bars are shown as they were of the same order of size as the symbols. The binocular visual field at any developmental stage is the overlap of the two monocular visual fields and is that area between the symbols and the naso-superior periphery of the visual field of the right eye.

Table 1 Degree of registration of ipsilateral and contralateral visuotectal projections to one tectum at different developmental ages

Age	Nasotemporal disparity (deg.)	Superoinferior disparity (deg.)
Stage 60-64 ($n=107$)	1.5 ± 12.6	3.2 ± 10.0
Stage 66 ($n=129$)	1.3 ± 10.0	0.3 ± 9.7
3MAM ($n=121$)	3.2 ± 10.8	0.2 ± 9.4
12MAM ($n=165$)	0.8 ± 12.8	0.5 ± 7.4
Adult ($n=137$)	0.3 ± 9.9	0.4 ± 7.9

Values are mean $\pm$ s.d. MAM, months after metamorphosis. n , Number of electrode penetrations.

60 to adulthood. Detailed maps permitting such analysis were obtained in animals at stage 60 (2), stage 62 (4), stage 64 (5), stage 66 (10), and post-metamorphically at 2 weeks (8), 6 weeks (2), 3 months (10), 1 yr (7) and adulthood (4). Figure 3 shows representative results from animals at stages 62 and 66 and 3 months after metamorphosis. We found that, at all developmental ages studied, the ipsilateral visuotectal projection arose only from the binocular portion of the visual field and effectively occupied the entire binocular visual space then existing. This projection thus increased in size with age as the binocular visual field enlarged. There was a corresponding increase in the proportion of the tectal surface occupied by the ipsilateral visuotectal projection.

Table 1 provides data on the degree of registration between the ipsilateral and contralateral visuotectal projections to one

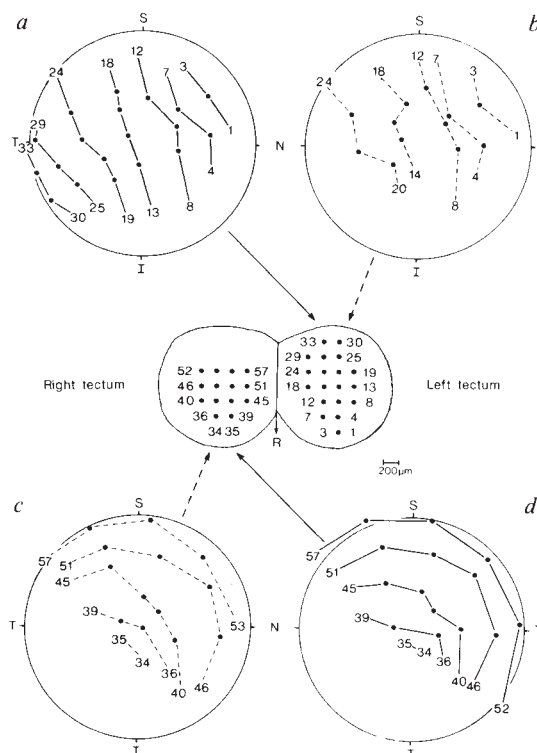


Fig. 2 The contralateral and ipsilateral visuotectal projections in a normal adult *Xenopus*. This animal, anaesthetized with MS 222 (ethyl *m*-aminobenzoate; Sigma) and subjected to partial craniotomy, was placed in an Aimark projection perimeter with its right eye centred (see Fig. 1 legend). The diagrams of the optic tecta show their outlines, transposed from a photograph of the dorsal surface, upon which a 100- μ m grid was superimposed. Rostral is indicated by the direction of the midline arrow labelled R. Numbers and intermediate dots on the tectal diagram represent the sites of vertical microelectrode penetrations. Large circles denote perimetric chart representations of the visual field, with the optic axis of the right eye centred on the fixation point of the perimeter as in Fig. 1. The animal was not moved within the coordinate frame during the recording session, thus the four chart representations plot identical areas of visual space. For each microelectrode position the centre of the region in visual space, stimulation of which produced evoked unitary potentials at that electrode site, when viewed by either the right or the left eye, is indicated by the corresponding number of intermediate dots on the relevant chart representation. *a*, The contralateral visuotectal projection through the right eye; *b*, the ipsilateral visuotectal projection through the left eye; *c*, ipsilateral visuotectal projection through the right eye; *d*, contralateral visuotectal projection through the left eye. The indirect nature of the ipsilateral visuotectal pathway is denoted by the discontinuous lines within and from the visual field representations. N, T, S and I are the nasal, temporal, superior and inferior aspects of the visual field of the tested eye. Note that the two visual projections to each tectum are in register, for example, equivalent visual field positions 35 in *c* and *d* project through both right and left eyes to microelectrode position 35 in the right tectum. The pattern of intertectal connections may be derived from the visuotectal projections as follows: right tectal position 35 is receiving visual input from visual field position 35 through the contralateral left eye. This visual field position is identical to field position 21 which projects through the left eye to left tectal position 21. The presence of an ipsilateral visuotectal response at the left tectal position 21 indicates that visual information, conveyed from the left eye directly to right tectal position 35, has been relayed through the intertectal system to left tectal position 21. When this type of analysis is extended to other parts of the visuotectal projections through the left eye, it yields the overall pattern of intertectal connections relaying information from the right to the left tectum. Similar analysis of the visuotectal projections through the right eye to both tecta reveals the pattern of intertectal connections from the left to the right tectum. The latter are, in fact, the reciprocal of those from the right to the left tectum.

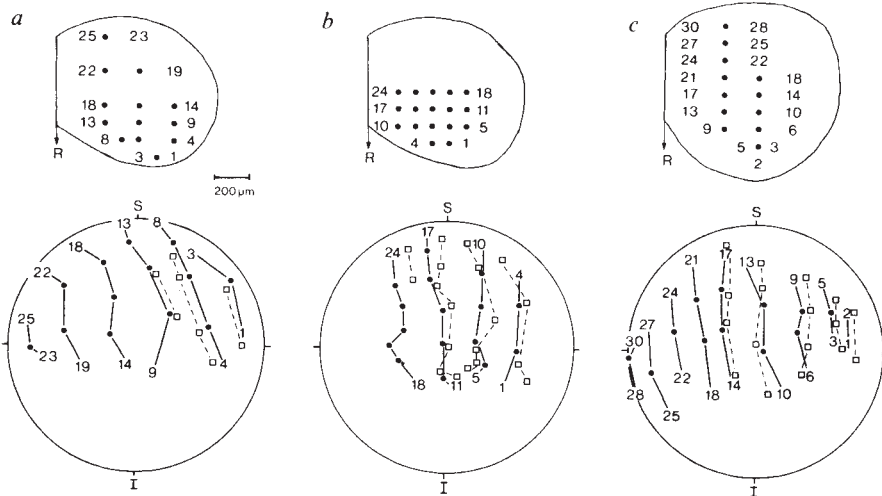


Fig. 3 Contralateral and ipsilateral visuotectal projections to the left tectum in representative animals at: *a*, stage 62; *b*, stage 66; and *c*, 3 months after metamorphosis. Conventions are essentially similar to those of Fig. 2, except that the two visuotectal projections are superimposed. The contralateral projections are represented by closed circles linked by continuous lines, the ipsilateral projections by open squares linked by dotted lines. Projections to the right optic tectum were also mapped but are not shown.

tectum at various developmental ages. The nasotemporal and superoinferior disparities between the two visual field positions to one tectal locus were measured and the results grouped by developmental age. Apparently, there were no systematic disparities at any developmental age, indicating that throughout development the ipsilateral visuotectal projection through one eye was in register with the contralateral visuotectal projection through the other eye.

The patterns of intertectal connections at various stages of development from stage 60 to adulthood were determined from the visuotectal projections by the method outlined in Fig. 2 legend. To compare the connections at different developmental stages, it is necessary to know the nature of tectal growth during this period. We have recently reported¹⁰ the results of a detailed study on histogenetic and hypertrophic growth of both the retina and the tectum from stage 60 to adulthood. Summarizing the tectal results briefly, there was no significant tectal histogenesis during this period. From stage 60 until 3 months after metamorphosis there was little change in the linear dimensions of the tectum although these did increase thereafter. Tectal cell density counts indicated that this increase in tectal size was uniformly hypertrophic; this information was used to construct the diagram in Fig. 4, which shows the different points on the left optic tectum to which a given site near the rostral pole of the right optic tectum projects, via the intertectal system, at different developmental stages. When the ipsilateral visuotectal projection first appears at stage 60, the rostral pole of the right tectum is linked to a rostral region in the left tectum. As development proceeds through metamorphosis and juvenile life into adulthood, this same area of the right tectum is linked progressively to more caudal left tectal areas. Similar results are obtained if we consider other points on the right tectum participating in the intertectal system. With time, progressively more caudal right tectal areas are incorporated into the system, these project initially to rostral left tectal areas and are then displaced more caudally on the left tectum as new input arrives from the right tectum. The intertectal system itself is not a direct link but involves a synaptic relay in the nucleus isthmi¹⁴⁻¹⁸. Our present data do not allow us to determine which component(s) of the tecto-isthmo-tectal pathway is involved in this normal developmental plasticity. Other studies^{17,19,20}, however, suggest that it is the crossed isthmo-tectal projection that is the plastic component.

We conclude that the normal maturation of the intertectal system in *Xenopus laevis* is a highly plastic process. In response to the changing functional requirements presented by changing interocular alignment, the intertectal system undergoes a continuous and systematic change in its functional connections.

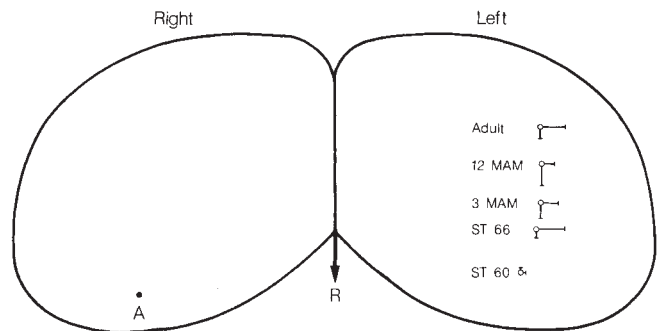


Fig. 4 Illustration of the changing intertectal connections during normal development in *Xenopus*. This diagram was constructed from patterns of intertectal connections derived, as explained in Fig. 2 legend, from animals of different developmental ages. For simplicity, the changing links of only one right tectal position (A) are shown. The diagram shows the positions on the left tectum which, at different stages of development, receive intertectal connections from the given position (A) near the rostral pole of the right tectum (MAM, months after metamorphosis; Ad, adult). As animals of different ages are involved in this analysis, allowance had to be made for tectal growth during this period (see text). The tectal diagrams represent tecta in adult animals. In plotting the results from younger animals on these tecta, the necessary corrections for tectal growth have been made. The open circles on the left tectum indicate the mean position for animals of that age (stage 60, $n = 2$; stage 66, $n = 10$; 3MAM, $n = 10$; 12MAM, $n = 7$; adult, $n = 4$). The bars indicate the standard deviation in the rostro-caudal and medio-lateral tectal axes, except in the case of the two stage-60 animals, where they indicate the range. The rostro-caudal tectal positions of all age groups shown are statistically different from each other. Intertectal connections from the rostral region of the right tectum project, with age, to progressively more caudal positions on the left tectum. The converse is also true. As the normal intertectal projection is reciprocally symmetrical, this diagram also shows that, with age, intertectal connections from the left tectum to the rostral region of the right tectum arise from progressively more caudal tectal positions. Calibration bar, 200 μ m.

The fact that the normal development of this system entails such continuous synaptic adjustments may explain why it is capable of responding with major readjustments to experimental perturbation of interocular alignment, as, for example, by early eye rotation^{17,19,21-23}. In both normal development and following eye rotation, the intertectal connections alter so as to link the new set of corresponding tectal points. Visual deprivation disrupts the normal development of the intertectal system^{20,24,25}.

and prevents the readjustment of intertectal connections following eye rotation²⁴.

In the cat the first few months of postnatal life see a maturation of eye alignment²⁶⁻³¹ involving maturational changes in cortical and callosal systems associated with binocular vision^{3,31-33}. These systems are also able to adjust following experimental alteration of eye alignment^{7,8,34-36}. It has been suggested that critical periods of experience-dependent development in both binocular visual³⁷⁻⁴⁰ and binaural auditory⁴¹ systems reflect a requirement for neuronal systems to adjust to growth-related changes in interocular or interaural geometry. In *Xenopus* the developmental changes in interocular geometry are particularly marked; the intertectal system in this species provides a clear demonstration that a neuronal system subserving binocular integration does, in fact, undergo extensive functional rewiring during normal maturation.

Received 11 November 1985; accepted 21 April 1986.

1. Redfern, P. A. *J. Physiol., Lond.* **209**, 701-709 (1970).
2. Crepel, F., Mariani, J. & Delhaye-Bouchaud, N. *J. Neurobiol.* **7**, 567-578 (1976).
3. Innocenti, G. M., Fiore, L. & Caminiti, R. *Neurosci. Lett.* **4**, 237-242 (1977).
4. LeVay, S., Wiesel, T. N. & Hubel, D. *J. comp. Neurol.* **191**, 1-51 (1980).
5. Raiman, G. *Brain Res.* **14**, 25-48 (1969).
6. Hubel, D., Wiesel, T. N. & LeVay, S. *Phil. Trans. R. Soc. B278*, 377-409 (1977).
7. Lund, R. D., Mitchell, D. E. & Henry, G. H. *Brain Res.* **144**, 169-172 (1978).
8. Berman, N. E. & Payne, B. R. *Brain Res.* **274**, 201-212 (1983).
9. Grobstein, P. & Comer, C. *Nature* **269**, 54-56 (1977).
10. Grant, S. & Keating, M. J. *J. Embryol. exp. Morph.* **92**, 43-69 (1986).

11. Nieuwenkoop, P. D. & Faber, J. *A Normal Table of Xenopus laevis* (Daudin) 2nd edn (North-Holland, Amsterdam, 1967).
12. Gaze, R. M. & Jacobson, M. Q. *Jl exp. Physiol.* **47**, 273-280 (1962).
13. Keating, M. J. & Gaze, R. M. Q. *Jl exp. Physiol.* **55**, 284-292 (1970).
14. Glasser, S. & Ingle, D. *Brain Res.* **159**, 214-218 (1978).
15. Grobstein, P., Comer, C., Hollyday, M. & Archer, S. M. *Brain Res.* **156**, 117-123 (1978).
16. Gruberg, E. R. & Udin, S. B. *J. comp. Neurol.* **179**, 487-500 (1978).
17. Udin, S. B. & Keating, M. J. *J. comp. Neurol.* **203**, 575-594 (1981).
18. Grobstein, P. & Comer, C. *J. comp. Neurol.* **217**, 54-74 (1983).
19. Udin, S. B. *Nature* **301**, 336-338 (1983).
20. Keating, M. J. & Kennard, C. *Neuroscience* (in the press).
21. Gaze, R. M., Keating, M. J., Szekely, G. & Beazley, L. *Proc. R. Soc. B175*, 107-147 (1970).
22. Keating, M. J. *Proc. R. Soc. B189*, 603-610 (1975).
23. Keating, M. J., Beazley, L., Feldman, J. D. & Gaze, R. M. *Proc. R. Soc. B191*, 445-466 (1975).
24. Keating, M. J. & Feldman, J. D. *Proc. R. Soc. B191*, 467-474 (1975).
25. Grant, S. & Keating, M. J. *J. Physiol., Lond.* **320**, 19P-20P (1981).
26. Sherman, S. M. *Brain Res.* **37**, 187-203 (1972).
27. Olsen, C. R. & Freeman, R. D. *Nature* **271**, 446-447 (1978).
28. Elberger, A. J. *Expl Brain Res.* **36**, 71-85 (1979).
29. von Grünau, M. W. *Expl Brain Res.* **37**, 41-47 (1979).
30. Pettigrew, J. D. *J. Physiol., Lond.* **237**, 49-74 (1974).
31. LeVay, S., Stryker, M. P. & Schatz, C. J. *J. comp. Neurol.* **179**, 223-244 (1978).
32. Innocenti, G. M. *Science* **212**, 824-827 (1981).
33. Hubel, D. & Wiesel, T. N. *J. Neurophysiol.* **28**, 1041-1059 (1965).
34. Schlaer, R. *Science* **173**, 638-641 (1971).
35. Innocenti, G. M. & Frost, D. O. *Nature* **280**, 231-234 (1979).
36. Dürsteler, M. R. & von der Heydt, R. *J. Physiol., Lond.* **345**, 87-105 (1983).
37. Keating, M. J. *Br. med. Bull.* **30**, 145-151 (1974).
38. Blakemore, C. in *Handbook of Sensory Physiology* Vol. 8 (eds Hedli, E., Liebowitz, M. W. & Teuber, H. L.) 377-436 (Springer, New York, 1978).
39. Pettigrew, J. D. in *Neuronal Plasticity* (ed. Cotman, C. W.) 311-330 (Raven, New York, 1978).
40. Mitchell, D. E. in *Development of Perception* Vol. 2 (eds Aslin, J. N., Alberts, J. R. & Peterson, M. R. 3-43 (Academic, New York, 1981).
41. Merzenich, M. M., Jenkins, W. M. & Middlebrooks, J. C. in *Dynamical Aspects of Neocortical Function* (eds Edelman, G. M., Gall, W. E. & Cowan, W. M.) 397-424 (Wiley, New York, 1984).

Measurement of the intracellular free calcium concentration in salamander rods

P. A. McNaughton, L. Cervetto* & B. J. Nunn

Physiological Laboratory, University of Cambridge, Downing St, Cambridge CB2 3EG, UK

Measurement of the free calcium concentration within a photoreceptor outer segment has been considered an important aim since the proposal by Hagins and Yoshikami^{1,2} that the primary event in phototransduction is a release of Ca^{2+} inside the cell. More recent evidence³⁻⁶ has cast doubt on the calcium hypothesis, and the observations of Yau and Nakatani⁷ and Matthews *et al.*⁶ suggest that the internal Ca^{2+} concentration ($[\text{Ca}^{2+}]_i$) may decrease after a flash of light. In the present study we have measured $[\text{Ca}^{2+}]_i$ directly by using a new method for incorporating the Ca-sensitive photoprotein aequorin into an isolated rod. We report that the light response is accompanied by a decrease in $[\text{Ca}^{2+}]_i$, caused by the closure of light-sensitive channels which are the main route for Ca^{2+} entry into the outer segment. Of the Ca^{2+} entering through light-sensitive channels, about 95% is sequestered by a rapid and reversible buffering mechanism. Calcium is removed from the cell by an electrogenic pump³ in which 3 Na^+ ions are exchanged for each Ca^{2+} ; the pump is highly active and the free Ca^{2+} in the cell declines with a time constant of ~ 0.5 s after a flash of light.

A rod mechanically isolated from the retina of the tiger salamander, *Ambystoma tigrinum*, was drawn, inner segment first, into a suction pipette, leaving the outer segment projecting into a solution which could be rapidly changed⁴. A whole-cell recording pipette was sealed against the outer segment membrane, and the membrane was broken by suction or by a pulse of current^{6,8,9}. Aequorin was then ejected from a fine plastic pipe inserted as close as possible (~ 100 μm) to the mouth of the whole-cell pipette, and was allowed to diffuse into the rod outer segment (ROS). The whole-cell pipette was gently withdrawn after ~ 10 min; the membrane usually reseals, leaving the aequorin-loaded rod in the suction pipette. Rods were loaded

with up to 4×10^4 counts using our method, corresponding to $\sim 4 \times 10^6$ aequorin molecules in the cell.

In normal Ringer's, the calcium concentration in the ROS is below the limit of resolution of the present method, and no significant change in $[\text{Ca}^{2+}]_i$ could be detected during the response to a flash of light. From the cell in which the best loading was obtained, an upper limit of 0.6 μM (at the 95% confidence level) could be set for the free calcium concentration both in the dark and at the plateau of the light response.

The calcium concentration can be increased to detectable levels by using the phosphodiesterase inhibitor IBMX (3-isobutyl-1-methylxanthine) which increases the intracellular level of cyclic GMP and consequently opens light-sensitive channels^{5,10,11}. Figure 1 shows that when IBMX was applied to the outer segment the light-sensitive current increased, and a burst of light was recorded from the aequorin-loaded cell. If the exposure to IBMX was prolonged the light-sensitive current oscillated, and each current maximum was accompanied by a burst of aequorin light emission. These results are difficult to reconcile with the idea that intracellular Ca^{2+} acts directly to close light-sensitive channels, since the light-sensitive current and free $[\text{Ca}^{2+}]_i$ increase at the same time. A more likely explanation is that light-sensitive channels are permeable to Ca^{2+} (for which there is other evidence^{3,4,12}) and that $[\text{Ca}^{2+}]_i$ rises because of the increase in light-sensitive current in the presence of IBMX. The oscillatory response to IBMX application was not seen in unloaded rods, and must be connected with absorption of the aequorin light by the photosensitive mechanism of the rod itself.

The effects of a flash of light on $[\text{Ca}^{2+}]_i$ were investigated in the experiment shown in Fig. 2. As in Fig. 1, $[\text{Ca}^{2+}]_i$ was first increased to measurable levels by applying IBMX, and a bright flash was then delivered at the peak of the increase in current (Fig. 2, trace 1). The flash caused a delayed decline in $[\text{Ca}^{2+}]_i$, and the free calcium level after the flash at no time exceeded that observed in the absence of a flash. A small, slowly declining inward current is visible at the plateau of the light response (marked with a triangle in trace 1 of Fig. 2a). This current has been attributed to the operation of a light-insensitive electrogenic Na/Ca exchange which reduces the Ca^{2+} level within the cell after a flash of light^{3,7,10}. Figure 2c supports this interpretation as the time course of the decrease in $[\text{Ca}^{2+}]_i$ correlates with the integral of this 'pumping current'.

* Permanent address: Istituto di Neurofisiologia del CNR, via San Zeno 51, 56100 Pisa, Italy.

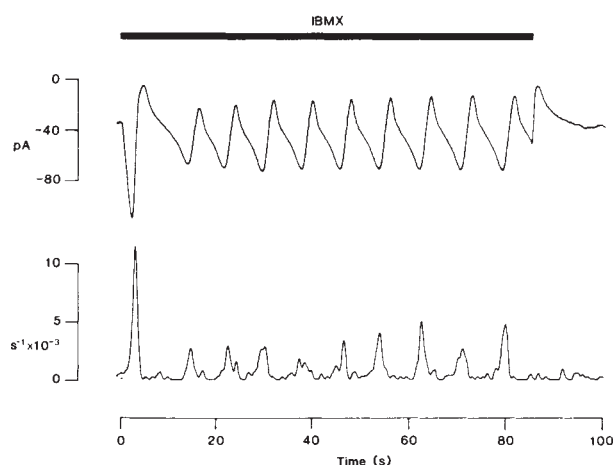


Fig. 1 Effect of exposure of the outer segment of an isolated rod loaded with aequorin to the phosphodiesterase inhibitor IBMX (0.5 mM). The upper panel shows the light-sensitive current recorded with a suction pipette. The lower panel shows the aequorin light output, expressed as rate of disintegration. The rate of disintegration was calculated as the ratio of the recorded count rate to the total counts in the cell, which were 5×10^3 ; the latter value was obtained by lysing the cell in 1 mM CaCl_2 at the end of the experiment and summing back the counts expended during the experiment. The period of oscillation was 8.3 s, and the peak aequorin light output occurred 0.63 s after the peak current.

Methods. The traces were low-pass-filtered at 20 Hz; the aequorin light signal was in addition filtered digitally by convolving with a gaussian function of standard deviation 275 ms. This filtering operation does not cause a temporal shift. Composition of the normal Ringer's was (in mM): NaCl 110, KCl 2.5, MgCl_2 1.6, CaCl_2 1, glucose, 3, HEPES 10, neutralized to pH 7.6 with TMAOH (tetramethylammonium hydroxide). Composition of the solution filling the whole-cell pipette was (in mM): K-aspartate 110, EGTA 0.02, MgCl_2 3, Na_2ATP 1, Na_2GTP 1, PIPES 10, neutralized to pH 7.2 with KOH. Aliquots of aequorin ($0.1 \mu\text{l}$ at 20 mg ml^{-1}) were dialysed against the filling solution before use. The aequorin light was collected with a $\times 100$ lens mounted on a Zeiss IM35 inverted microscope, and was focused onto the cathode of a photomultiplier tube (EMI 9789B) mounted in place of the binocular head. Single photons were counted by a Brookdeal quantum photometer. An aperture mounted at an intermediate image plane limited light collection to an area slightly greater than that of the outer segment. Aequorin light came principally from the outer segment, as little light was recorded with the aperture positioned over the inner segment¹³. The aperture was removed before lysis of the cell at the end of the experiment.

The influx of Ca^{2+} into the cell can be measured directly by exposing the ROS to a solution in which all permeant ions have been replaced by Ca^{2+} . Figure 3 shows an experiment of this kind. The influx of Ca^{2+} on exposure to isotonic Ca^{2+} was followed by a rise in the emission of light from the aequorin and a subsequent closure of light-sensitive channels. In contrast to the experiment of Fig. 2 (in the presence of a normal concentration of Na), there was little decline in the light emission after the closure of the channels; the observed slow decline in Fig. 3c is attributable largely to consumption of aequorin by the large increase in $[\text{Ca}^{2+}]_i$. On restoration of a normal concentration of Na, there was a rapid decline in $[\text{Ca}^{2+}]_i$, and a large light-insensitive pumping current. The rate of decline of free $[\text{Ca}^{2+}]_i$ on restoration of Na was $\sim 30 \mu\text{M s}^{-1}$, compared with a rate of $< 0.6 \mu\text{M s}^{-1}$ in the absence of Na (see Fig. 3 legend). This experiment shows that the maximal activity of the Na/Ca exchange is at least 50 times greater than that of other Ca pumps in the outer segment membrane.

In the experiment shown in Fig. 3, the total charge flowing during the application of isotonic Ca^{2+} was 46 pC and the total charge transferred by the pump was 20.6 pC, consistent with a $3 \text{ Na}^+ / 1 \text{ Ca}^{2+}$ stoichiometry for the pump³. In this experiment

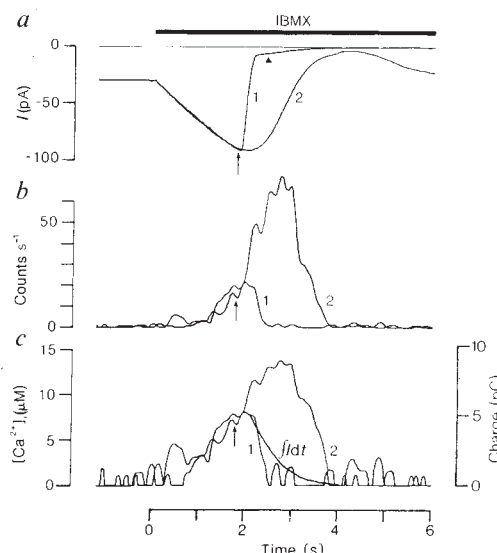


Fig. 2 Effect of a flash of light on $[\text{Ca}^{2+}]_i$. *a*, Light-sensitive current obtained with a light flash ($4.0 \times 10^4 \text{ Rh}^*$) delivered at the arrow (trace 1), or in the dark (trace 2). The triangle marks the pumping current at the plateau of the light response. *b*, The aequorin light output. *c*, $[\text{Ca}^{2+}]_i$ calculated from the rate of disintegration of aequorin by using a calibration curve supplied by Dr J. Blinks. Over the range $0.1 \mu\text{M} < [\text{Ca}^{2+}] < 30 \mu\text{M}$ and with free $[\text{Mg}^{2+}] = 1 \text{ mM}$, the curve was well fitted by the function $[\text{Ca}^{2+}] = A(\text{rate of disintegration})^y$, where $A = 4.67 \times 10^{-5} \text{ M}$ and $y = 0.417$. The integral of the pumping current (smooth trace, right-hand ordinate) is also shown in *c*. Traces are the mean of five presentations. The gaussian filter for the aequorin trace had a standard deviation of 65 ms. Total counts in the cell were 5.3×10^3 , and we assumed for the purpose of calculating $[\text{Ca}^{2+}]_i$ that 26% of the aequorin was in the outer segment since this was the maximum fraction of the total counts which could be exhausted by a rapid elevation of $[\text{Ca}^{2+}]_i$.

the suction pipette collected 52% of the outer segment current, as judged by the ratio of bright flash responses recorded by the whole-cell pipette to those recorded by the suction pipette at the start of the experiment. From this collection ratio, and taking an outer segment volume of 2 pl, of which half is assumed to be occupied by the disks, we calculate that the total $[\text{Ca}^{2+}]_i$ increased by $458 \mu\text{M}$ in the cytoplasmic space of the outer segment; the observed rise in $[\text{Ca}^{2+}]_i$ was less than this value by a factor of ~ 20 , showing that about 95% of the Ca^{2+} entering the outer segment is bound and is not free to interact with aequorin. The binding mechanism must be rapid and reversible, as there was a good temporal correlation between the change in $[\text{Ca}^{2+}]_i$, as measured from the aequorin light output, and the integrals of both the Ca^{2+} influx and the pumping current.

Figure 4 shows the results of an experiment in which a large increase in light-sensitive current resulted from exposure to low external Ca^{2+} (ref. 4) in the absence of IBMX. In this experiment the increase in current did not cause an increase in $[\text{Ca}]_i$, but on return to Ringer's containing 1 mM Ca there was a large influx of calcium through light-sensitive channels and an increase in aequorin light emission. The subsequent decline in $[\text{Ca}^{2+}]_i$ correlates well with the time course of charge transfer by the pump (see Fig. 4c). The rate of decline was slower than in the experiments of Figs 2 and 3, perhaps because of the build-up of Na_i during the preceding large current flow. From the total charge transferred by the pump we calculate that 24% of the current flowing on return to normal Ringer's was carried by Ca^{2+} .

The results described here can be used to estimate the normal free $[\text{Ca}^{2+}]_i$ when the rod is bathed in Ringer's. In the presence of a bright light $[\text{Ca}^{2+}]_i$ is likely to be low, in view of our finding that the main route of entry is via the light-sensitive channel.

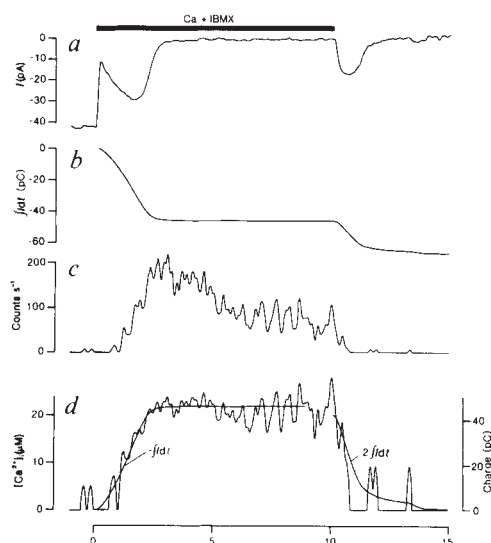


Fig. 3 *a*, Membrane current recorded during exposure to an isotonic CaCl_2 solution containing (in mM): CaCl_2 77.6, HEPES 10 (pH 7.6), IBMX 0.5. The junction current has been subtracted⁴. The elevation in Ca^{2+} caused an initial rapid suppression of current, due to an external action⁴, followed by a rise due to the inhibition of phosphodiesterase by the IBMX. The subsequent decrease is due to closure of light-sensitive channels by the aequorin light emission; the light-sensitive current remained completely suppressed until after the end of the trace shown in *a*. *b*, Integral of the membrane current in *a*, beginning from the time of admission of isotonic Ca^{2+} . Note that the integral of the current carried by Ca^{2+} is approximately double that of the pumping current. *c*, The aequorin light output. *d*, Calculated change in $[\text{Ca}^{2+}]_i$ (noisy trace, left-hand ordinate) based on the assumptions outlined in Fig. 2 legend. Based on these assumptions, no decline in $[\text{Ca}^{2+}]_i$ is observed during exposure to isotonic Ca^{2+} , and the rate of decline on restoration of Na is $30 \mu\text{M s}^{-1}$. If instead we make the extreme assumption that all the aequorin is in the outer segment, then the rate of decline of $[\text{Ca}^{2+}]_i$ is $0.6 \mu\text{M s}^{-1}$ during the exposure to isotonic Ca^{2+} . Total counts in the cell were 4.6×10^3 . The smooth traces (right-hand ordinate) in *d* compare the time courses of the integrals of the light-sensitive and pumping currents with the time course of the change in $[\text{Ca}^{2+}]_i$.

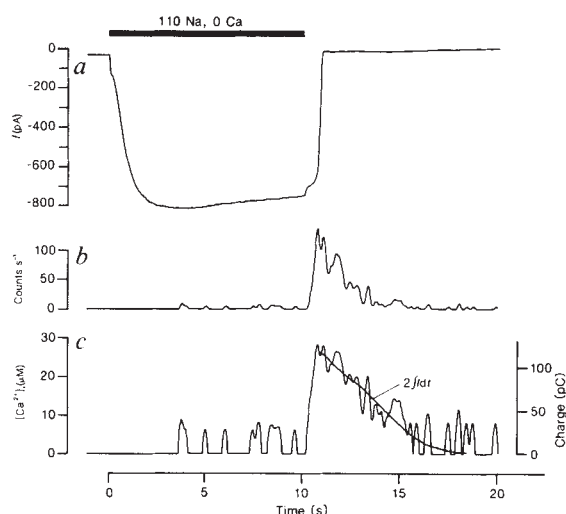


Fig. 4 Membrane current (*a*), aequorin light output (*b*) and $[\text{Ca}^{2+}]_i$ (*c*) during exposure to Ringer's containing 0 Ca and 2 mM BAPTA¹⁴. The smooth trace (right-hand ordinate) in *c* is the integral of the pumping current observed after all light-sensitive channels had been closed by the strong aequorin light emission. The absence of bumps at the beginning of the traces in *b* and *c*, which was not seen in other records, is not significant. Total counts in the cell, 1.86×10^3 . For other details, see Fig. 2 legend.

After a bright flash of light the pumping current at the plateau of the response declines with a time constant of ~ 0.5 s, and carries a total charge of ~ 0.5 pC, as measured with a suction pipette^{7,10}. If we assume a current collection efficiency of 50% and if the pump operates as a $3\text{Na}^+/1\text{Ca}^{2+}$ exchange, this charge transfer corresponds to a change in total $[\text{Ca}^{2+}]_i$ of $10 \mu\text{M}$ when distributed over an effective outer segment volume of 1 pl. If 95% of this Ca^{2+} is bound, the free $[\text{Ca}^{2+}]_i$ in the dark will be $0.5 \mu\text{M}$, which is slightly below our present detection limit of $0.6 \mu\text{M}$. As we find that the time course of decline of free $[\text{Ca}^{2+}]_i$ correlates well with the time course of the integral of the pumping current, we expect that the effect of a bright flash will be to reduce $[\text{Ca}^{2+}]_i$ from its dark level of $\sim 0.5 \mu\text{M}$ to a much lower level with a time constant of ~ 0.5 s.

This work was supported by the MRC and by an EMBO fellowship (to L.C.). B.J.N. is a Lister Institute Research Fellow. We thank Drs J. R. Blinks and P. H. Cobbold for supplies of aequorin and for advice, and Professor A. L. Hodgkin, Dr T. D. Lamb and Mr L. Lagnado for helpful discussions.

Received 12 December; accepted 17 April 1986.

- Hagins, W. A. *Rev. Biophys. Bioeng.* **1**, 131-158 (1972).
- Hagins, W. A. & Yoshikami, S. *Expl Eye Res.* **18**, 299-305 (1974).
- Yau, K. W. & Nakatani, K. *Nature* **311**, 661-663 (1984).
- Hodgkin, A. L., McNaughton, P. A. & Nunn, B. J. *J. Physiol., Lond.* **358**, 447-468 (1985).
- Fesenko, E. E., Kolesnikov, S. S. & Lyubarsky, A. L. *Nature* **313**, 310-313 (1985).
- Matthews, H. R., Torre, V. & Lamb, T. D. *Nature* **313**, 582-585 (1985).
- Yau, K.-W. & Nakatani, K. *Nature* **313**, 579-582 (1985).
- Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. *Eur. J. Physiol.* **391**, 85-100 (1981).
- Cobbs, W. H. & Pugh, E. N. *Nature* **313**, 585-587 (1985).
- Cervetto, L. & McNaughton, P. A. *J. Physiol., Lond.* **370**, 91-109 (1986).
- Capovilla, M., Caretta, A., Cavaggioni, A., Cervetto, L. & Sorbi, R. T. *Prog. Retinal Res.* **2**, 233-247 (1983).
- Capovilla, M., Caretta, A., Cervetto, L. & Torre, V. *J. Physiol., Lond.* **343**, 295-310 (1983).
- Cervetto, L., McNaughton, P. A. & Nunn, B. J. *J. Physiol., Lond.* **371**, 36P (1986).
- Tsien, R. Y. *Biochemistry* **19**, 2396-2404 (1980).

Mediation of thalamic sensory input by both NMDA receptors and non-NMDA receptors

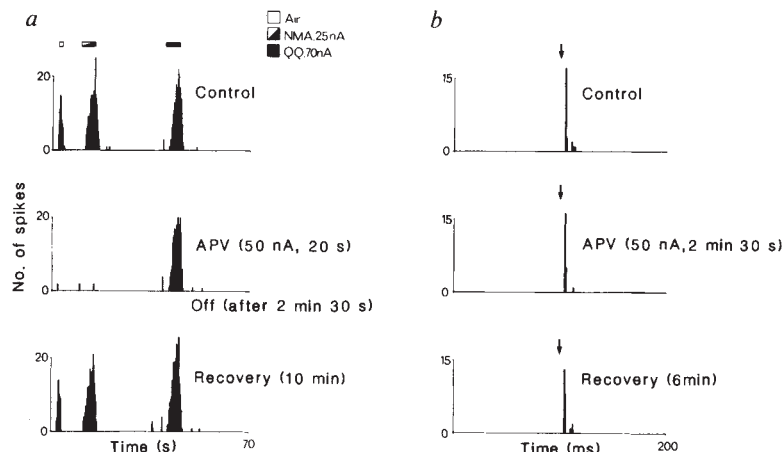
T. E. Salt

Department of Physiology, University College, Cardiff, PO Box 78, Cardiff CF1 1XL, UK

Excitatory amino acids such as L-glutamate and L-aspartate are well established as neurotransmitter candidates in the mammalian central nervous system¹, and three types of receptor for these substances have been proposed, characterized by the agonists N-methyl-D-aspartate (NMDA), kainate and quisqualate¹. All these receptors have been suggested to have synaptic roles in excitatory transmission in the brain¹⁻⁷. Here I demonstrate that NMDA receptors play a crucial role in the observed response of ventrobasal thalamus (VB) neurones to natural stimulation of somatosensory afferents, but do not appear to be responsible for the short-latency excitation seen on electrical stimulation of the afferents, which is apparently mediated by excitatory amino-acid receptors of the non-NMDA type. This result indicates an involvement of both NMDA and non-NMDA receptors in the responses of VB neurones to stimulation of somatosensory afferents, depending on the mode of stimulation of the pathway.

Recordings, using extracellular iontophoretic electrodes, were made from single neurones in the ventrobasal thalamus of rats anaesthetized with urethane (1.2 g per kg intraperitoneal). Responses of VB neurones to physiological stimulation of hair or vibrissa follicle afferents (using an air jet) and iontophoretic application of N-methylaspartate (NMA) and either kainate (14 neurones) or quisqualate (14 neurones), were challenged

Fig. 1 Peristimulus time histograms of action potentials recorded from a single VB neurone with one barrel of a five-barrelled iontophoretic electrode. Action potentials were discriminated on the basis of amplitude and waveform. The vertical axis of each histogram is the number of action potential spikes counted into epochs of either 250 ms (*a*) or 1 ms (*b*). Each drug barrel of the electrode contained one of the following: *N*-methylaspartate (NMA, 0.2 M, pH 8.0), quisqualate (QQ, 25 mM in 75 mM NaCl, pH 8.0), kainate (KA, 0.1 M, pH 8.0), APV (50 mM in 150 mM NaCl, pH 8.0), kynurenate (KYN, 50 mM, pH 8.0), acetylcholine (ACh, 0.5 M, pH 3.5), carbamylcholine (CCh, 0.2 M, pH 4.0) or 1 M NaCl. Automatic current balancing was used in this experiment, but qualitatively identical results were found in experiments where balancing was not performed. In three experiments a seven-barrelled electrode was used. *a*, Top record shows the control responses (averaged over two trials) of a neurone to an air jet, directed so as to deflect the hairs within its receptive field, and to iontophoretic application of NMA and QQ. The three stimuli were applied at the time and for the durations indicated by the bars above the record. The middle record shows the responses of the same neurone to the same three stimuli, but during the concurrent ejection of APV. APV antagonized the responses to the air jet stimulus and NMA, but had little effect on the response to QQ. The bottom record shows the same sequence, but 10 min after termination of the APV ejection when responses to the air jet and NMA had recovered to their control levels. *b*, Response of the same neurone as in *a* to electrical stimulation (0.1-ms pulse, 10 V, 0.5-Hz repetition rate) of the afferent pathway delivered via needle electrodes within the peripheral receptive field. In all experiments the stimulus intensity was adjusted to the minimum necessary to evoke a short-latency response. The time of stimulation is indicated by an arrow above each histogram. The records are averaged over 20 trials. The middle record shows the response of the neurone to the stimulus during the ejection of APV. The antagonist had little effect on the electrically evoked response, even though the APV ejection was for a longer duration than that necessary to antagonize responses to the air jet and NMA.



with concurrent application (0–50 nA) of the selective NMDA antagonist D-2-amino-5-phosphonovalerate (APV)¹. This antagonist was found to antagonize the responses to NMA but not responses to kainate or quisqualate, and had little effect on the spontaneous activity of VB neurones. It was also possible to antagonize the responses of the VB neurones to the physiological stimulus by 42–100% (mean = 80%), using the same APV ejection currents, in 26 of the 28 neurones studied in this way (Fig. 1*a*). The response of 11 of these neurones to single-pulse electrical stimulation of the afferent pathway was then studied. The

short-latency (4–10 ms) response evoked by such a stimulus was challenged with the same iontophoretic current of APV that had previously been found to be effective against NMA and the physiologically evoked response, and in all of these cases the antagonist was found to be ineffective against the electrically evoked response (Fig. 1*b*). In contrast, the broad-spectrum excitatory amino-acid antagonist kynurenate^{4,8} (10–150 nA), which antagonized the responses of VB neurones to quisqualate in addition to the responses to NMA and the physiological stimulus, blocked the responses to electrical stimulation by 54–100% (mean = 86%) in 11 neurones (Fig. 2). Responses to the cholinergic agonists acetylcholine and carbamylcholine remained largely unaffected by kynurenate, although in some cases a small excitatory effect of the antagonist became evident.

These results suggest that an excitatory amino acid functions as the transmitter mediating the responses of VB neurones to stimulation of second-order sensory afferents. This mechanism is consistent with those suggested previously^{9,10}, and with data on the retino-geniculate projection^{11,12}. In addition, it seems likely that the excitation observed in response to stimulation of the afferent pathway is composed of at least two separate components. The first of these components, which can be selectively evoked at short latency by single-pulse electrical stimulation in the peripheral receptive field, is apparently mediated by excitatory amino-acid receptors of the non-NMDA type, as it is antagonized by kynurenate but not by APV. It is unlikely that this response is resistant to APV because the response has a high safety factor, as threshold stimulus intensities were used and the response could be antagonized by kynurenate, which might be expected to be ineffective if the response had a high safety factor. The finding that the response of VB neurones to physiological stimulation of the sensory afferents is greatly reduced by the selective NMDA antagonist APV implies an additional response component, which is mediated by NMDA receptors and which is recruited during physiological stimulation. It is tempting to speculate that it is the pattern of afferent activity which governs the recruitment of the APV-sensitive response, as the response of the VB neurones to the air jet is a train of action potentials (typically of frequency ~20 Hz) lasting for the duration of the stimulus (0.5–2 s), while the response to the single-pulse electrical stimulus is one or two action potentials. To test this possibility further, electrical stimulation was

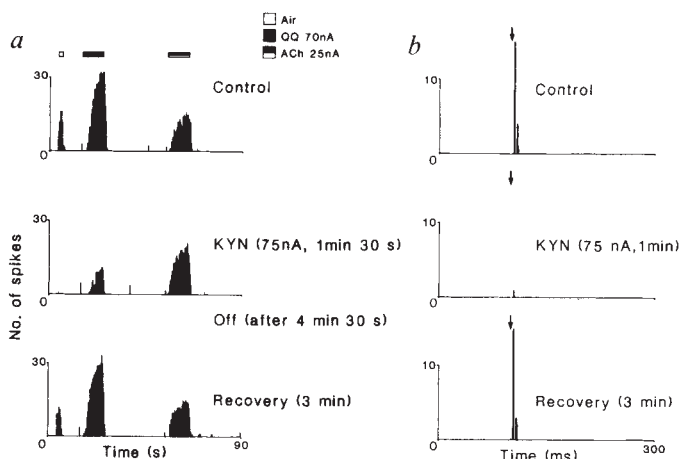


Fig. 2 *a*, Similar records to those in Fig. 1, but from another neurone which responded to the air jet stimulus and to iontophoretic ejection of QQ and ACh applied as indicated by the bars above the records. Below the control record is shown the same sequence during the ejection of kynurenate (KYN). A complete antagonism of the air jet response was evident, and the responses to QQ were also greatly reduced. No reduction in the response to ACh was seen. The bottom record shows recovery from the effects of kynurenate 3 min after the end of the antagonist ejection. *b*, Responses of the same neurone to single-pulse electrical stimulation in its peripheral receptive field at the time indicated by the arrows (15 trials). Middle record shows the response of the neurone during the ejection of kynurenate. The antagonist greatly reduced the response to the stimulus.

performed in 20-Hz trains on 10 neurones in which APV was found to antagonize the physiologically evoked response. For nine of these neurones, APV was found to antagonize the response to the stimulus train. In particular, the later components of the response appeared more susceptible to APV (Fig. 3). Note that the APV-sensitive component of this response does not appear to be larger than the APV-insensitive component; this suggests that the APV-insensitive component becomes less effective during the stimulus train, possibly because of activation of inhibitory processes in the thalamus¹³. Clearly, experiments using intracellular recording would provide further information on this matter.

It is conceivable that, under physiological conditions, the NMDA receptor component potentiates an already present, but possibly sub-threshold, excitatory input mediated by non-NMDA excitatory amino-acid receptors (presumably of the kainate or quisqualate type). This hypothesis is attractive in view of recent work which shows that responses of neurones to NMDA are subject to voltage-dependent blockade by Mg^{2+} ions in the physiological concentration range for Mg^{2+} (refs 6, 14, 15), and it is noteworthy that a postsynaptic potential involving a short-latency non-NMDA component and an NMDA component of longer latency have been found in the spinal cord of *Xenopus*¹⁶. The recruitment of NMDA receptors could be a purely synaptic phenomenon similar to the APV-sensitive long-term potentiation observed in the hippocampus⁷, and might involve one type of excitatory amino acid being released onto both NMDA and non-NMDA receptors. Alternatively, the two types of response could be mediated by two afferent pathways to the thalamus which terminate on the same neurone. Evidence for such multiple pathways already exists^{17,18}.

The NMDA receptor-mediated response of VB neurones to physiological stimulation of afferents provides a possible site of action for the dissociative anaesthetic ketamine, which has been shown to possess NMDA antagonist properties^{6,19}. This contrasts with previous findings in the dorsal horns of the spinal cord²⁰ and medulla³, where NMDA antagonists were found to be ineffective in blocking responses to air jet deflection of hairs and vibrissae. As there are no major procedural differences between the present experiments and those performed on the medullary dorsal horn³, it seems likely that the observed difference reflects a regional difference between the dorsal horn and thalamus. Thus, it is possible that the thalamus is a major site for the anaesthetic action of ketamine.

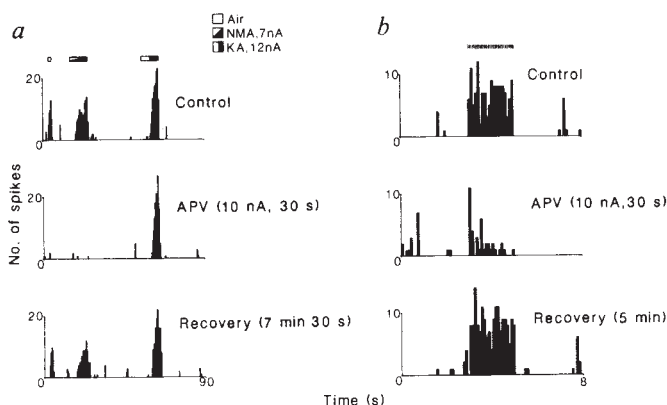


Fig. 3 *a*, Similar records to those in Fig. 1 (500 ms epochs), but from a neurone responding to air jet stimulation and iontophoretically applied NMA and kainate (KA). APV selectively and reversibly antagonized the responses to the air jet and NMA. *b*, Records from the same neurone as in *a*, showing the responses to a train of electrical stimuli (20 Hz, 0.1-ms pulse, 2 s duration, 100 ms epochs, 5 trials). The time of stimulation is indicated by the vertical lines above the top record. The middle record shows the response of the neurone during the ejection of APV, and the bottom record shows the recovery after termination of the APV ejection.

I thank Ms V. Crimmins for technical assistance. This work was supported by the MRC (UK).

Received 14 January; accepted 16 April 1986.

- Watkins, J. C. & Evans, R. H. *A. Rev. Pharmac. Tox.* **21**, 165-204 (1981).
- Davies, J. & Watkins, J. C. *Brain Res.* **327**, 113-120 (1985).
- Hill, R. G. & Salt, T. E. *J. Physiol., Lond.* **327**, 65-78 (1982).
- Herrling, P. L. *Neuroscience* **14**, 417-426 (1985).
- Crunelli, V., Fonda, S. & Kelly, J. S. *J. Physiol., Lond.* **341**, 627-640 (1983).
- Thomson, A. M., West, D. C. & Lodge, D. *Nature* **313**, 479-481 (1985).
- Collingridge, G. L., Kehrl, S. J. & McLennan, H. *J. Physiol., Lond.* **334**, 33-46 (1983).
- Perkins, M. N. & Stone, T. W. *Brain Res.* **274**, 184-187 (1982).
- Halldeman, S., Huffman, R. D., Marshall, K. C. & McLennan, H. *Brain Res.* **39**, 419-425 (1972).
- Ben-Ari, Y. & Kelly, J. S. *J. Physiol., Lond.* **251**, 25P-27P (1975).
- Crunelli, V., Leresche, N. & Pirchio, M. *J. Physiol., Lond.* **365**, 40P (1985).
- Kemp, J. A. & Sillito, A. M. *J. Physiol., Lond.* **323**, 377-391 (1982).
- Andersen, P., Eccles, J. C. & Sears, T. J. *J. Physiol., Lond.* **174**, 370-399 (1964).
- Nowak, L., Bregestovski, P., Ascher, P., Herbert, A. & Prochiantz, A. *Nature* **307**, 462-465 (1984).
- Mayer, M. L. & Westbrook, G. L. *J. Physiol., Lond.* **361**, 65-90 (1985).
- Dale, N. & Roberts, A. J. *J. Physiol., Lond.* **363**, 35-59 (1985).
- Hongo, T. & Koike, H. in *The Somatosensory System* (ed. Kornhuber, H. H.) 218-226 (Georg Thieme, Stuttgart, 1975).
- Albe-Fessard, D., Berkley, K. J., Kruger, L., Ralston, H. J. III & Willis, W. D. *Brain Res. Rev.* **9**, 217-296 (1985).
- Anis, N. A., Berry, S. C., Burton, N. R. & Lodge, D. *Br. J. Pharmac.* **79**, 565-575 (1983).
- Anis, N. A., Headley, P. M., Lodge, D. & West, D. C. *J. Physiol., Lond.* **328**, 10P (1982).

Frequency-dependent involvement of NMDA receptors in the hippocampus: a novel synaptic mechanism

Caroline E. Herron, Robin A. J. Lester,
Elizabeth J. Coan & Graham L. Collingridge

Department of Pharmacology, University of Bristol, Medical School,
University Walk, Bristol BS8 1TD, UK

Acidic amino acids, such as L-glutamate, are believed to be excitatory neurotransmitters in the mammalian brain^{1,2} and exert effects on several different receptors named after the selective agonists kainate, quisqualate and *N*-methyl-D-aspartate (NMDA)¹. The first two receptors, collectively termed non-NMDA receptors, have been implicated in the mediation of synaptic transmission in many excitatory pathways in the central nervous system (CNS), whereas NMDA receptors, with few exceptions, do not appear to be involved²; this is typified in the hippocampus where there is a high density of NMDA receptors³ yet selective NMDA receptor antagonists, such as D-2-amino-5-phosphonovalerate (APV), do not affect synaptic potentials⁴⁻¹¹. NMDA receptors have, however, been shown to be involved in long-term potentiation (LTP) in the hippocampus⁶⁻¹¹, a form of synaptic plasticity¹² which may be involved in learning and memory¹¹. NMDA receptors have also been found to contribute to epileptiform activity in this region^{13,14}. We now describe how NMDA receptors can participate during high-frequency synaptic transmission in the hippocampus, their involvement during low-frequency transmission being greatly suppressed by Mg^{2+} . A frequency-dependent alleviation of this blockade provides a novel synaptic mechanism whereby a single neurotransmitter can transmit very different information depending on the temporal nature of the input. This mechanism could account for the involvement of NMDA receptors in the initiation of LTP and their contribution, in part, to epileptic activity.

Experiments were performed on transverse hippocampal slices prepared from adult female rats as described previously¹⁵. The standard perfusion medium comprised (in mM): NaCl 124, NaHCO₃ 26, NaH₂PO₄ 1.2, KCl 3 or 5, MgSO₄ 1 or 2, CaCl₂ 2, D-glucose 10. Extracellular and intracellular recordings were obtained from the CA1 cell body region using electrodes containing NaCl (1-4 M) and CH₃COOK (3 M), respectively. The Schaffer collateral-commissural pathway was stimulated with single shocks delivered at 10-30-s intervals to elicit 'low-

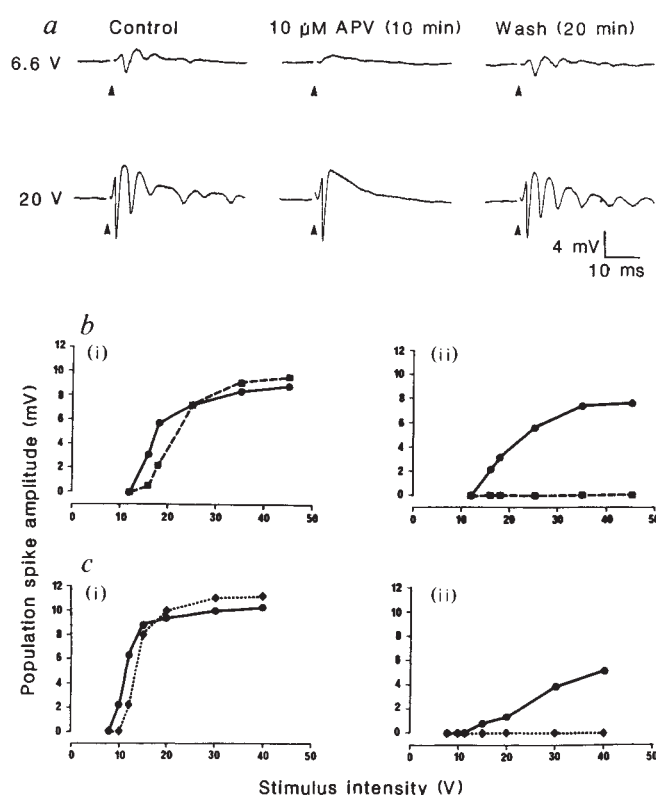


Fig. 1 Effects of NMDA antagonists on field potentials recorded extracellularly from the CA1 cell body region in response to low-frequency stimulation of the Schaffer collateral-commissural pathway. *a*, Effect of 10 μM APV on responses at two stimulus intensities (6.6 V and 20 V). *b*, Amplitude of the primary (i) and secondary (ii) population spikes as a function of stimulus intensity in control medium (continuous line) and in the presence of 20 μM APV (broken line) in another hippocampal slice. *c*, Data presented in the same manner as *b* for a different slice, showing the effects of 100 μM Mg²⁺.

frequency' synaptic responses and for periods of 200 ms at 10-ms intervals to evoke 'high-frequency' responses. The selective NMDA antagonist APV was administered via the perfusion medium (flow rate 1–2 ml min⁻¹). In the absence of APV, high-frequency stimulation often produced LTP of the excitatory postsynaptic potential (e.p.s.p.), therefore periods of high-frequency stimulation were delivered first in the presence and subsequently following washout of APV. The data were tested for significance ($P < 0.05$) using paired Student's *t*-tests.

Field potentials were recorded from the CA1 cell body region in response to low-frequency stimulation of the Schaffer collateral-commissural pathway with no Mg²⁺ in the bathing medium¹⁶. At low stimulus intensities the synaptic response was greatly suppressed (Fig. 1*a*) or abolished by 10 or 20 μM APV. In contrast, at higher stimulus intensities, the secondary population spikes were abolished while the primary population spike was little affected by APV (Fig. 1*a, b*), presumably because non-NMDA excitatory amino-acid receptors could sustain this component of the response⁶. These effects were fully reversible and seen in all 10 hippocampal slices examined. An identical pattern of antagonism was also observed with Mg²⁺ (100 μM) in all four slices tested (Fig. 1*c*).

These results demonstrate for the first time that NMDA receptors mediate a synaptic response in hippocampal neurones that has as low a threshold as that of non-NMDA receptors, provided Mg²⁺ is absent from the extracellular medium; this could be accounted for by the neurotransmitter having a high affinity for NMDA receptors, which would only be seen functionally when the Mg²⁺ block of NMDA channels was removed. The con-

centrations of Mg²⁺ (dose-dependent over the range 20–500 μM) that affect this component are the same as those that directly block postsynaptic responses to NMDA¹ and are much lower than those expected to affect significantly transmitter release¹⁷ or membrane stabilization¹⁸; they are also much lower than the generally accepted levels of Mg²⁺ in extracellular fluid. It is likely, therefore, that a major action of Mg²⁺ in the hippocampus is to exert a powerful controlling influence on synaptic transmission by a direct interaction with the NMDA receptor system.

A second series of experiments was performed to determine the extent of the involvement of the NMDA receptor system in synaptic transmission in the hippocampus in the presence of physiological concentrations of Mg²⁺ (1–2 mM). Passive membrane properties, determined for 27 CA1 neurones (resting membrane potential -69 ± 1 mV; input resistance, calculated near the resting membrane potential, 27 ± 1 MΩ), were not altered significantly by APV (20 or 50 μM). Stimulation of the Schaffer collateral-commissural pathway at low frequencies evoked a fast e.p.s.p. which was unaffected by APV (Fig. 2*a*). Measurements (mean $\pm$ s.e.m.) of peak amplitude and duration at 50% of peak amplitude in control solutions (8.3 ± 0.6 mV; 9.6 ± 0.8 ms) and in the presence of APV (8.4 ± 0.6 mV; 9.3 ± 0.6 ms) were not significantly different ($n = 32$). The inhibitory postsynaptic potential (i.p.s.p.) which followed the fast e.p.s.p. was generally unaffected by APV (Fig. 2*a*).

In contrast to the lack of effect on low-frequency synaptic responses, APV had a marked effect on the synaptic response generated by high-frequency stimulation. During high-frequency stimulation, fast e.p.s.ps were associated with a slow depolarizing potential. APV (20 or 50 μM) had no effect on fast e.p.s.ps evoked during high-frequency stimulation (Fig. 2*c*); it did, however, reduce the size of the associated slow depolarization, measured from the pre-stimulus membrane potential to the base of the fast e.p.s.ps, in 17 of 22 neurones examined (Fig. 2*b, c*). The mean ($\pm$ s.e.m.) amplitude of the depolarization, measured at 140–150 ms into the high-frequency train, for these 17 cells in the presence and after washout of APV, was 5.3 ± 0.9 mV and 8.3 ± 1.0 mV, respectively. The size of the NMDA receptor-mediated potential calculated individually for each cell varied between 0.9 and 5.1 mV (mean $\pm$ s.e.m., 3.0 ± 0.2 mV). This NMDA receptor-mediated e.p.s.p. became apparent in some cells as early as the second fast e.p.s.p. in the high-frequency train (Fig. 2*b, c*).

The lack of effect of APV on low-frequency synaptic transmission in the Schaffer collateral-commissural pathway is in agreement with previous reports using extracellular recording^{5–9}, but contrasts with a recent intracellular study which reported that APV depressed the fast e.p.s.p.¹⁹. The latter effect can be attributed to the use of high concentrations of the racemic mixture of APV, which has previously been shown to depress synaptic transmission in this pathway by an action unrelated to NMDA-receptor antagonism⁷. As low-frequency synaptic responses in this pathway can be reduced by broad-spectrum excitatory amino-acid antagonists, such as γ -D-glutamylglycine⁶, it is believed that receptors of the kainate or quisqualate (that is, non-NMDA) type are responsible for fast e.p.s.ps. The present study has shown that an NMDA receptor-mediated e.p.s.p. can be recorded in this pathway during high-frequency stimulation, but the slow nature of this synaptic response as well as its activation characteristics clearly distinguished it from the fast e.p.s.ps. The observation that the NMDA receptor-mediated synaptic potential could often be detected as early as the second fast e.p.s.p. during high-frequency stimulation (see Fig. 2*b, c*) demonstrates that as few as two appropriately timed synaptic inputs may activate the NMDA receptor system in the hippocampus.

The finding that in the absence of Mg²⁺, low-frequency stimulation activates a low-threshold NMDA receptor-mediated synaptic response, implies that NMDA receptors are normally acted on by the neurotransmitter released by low-frequency

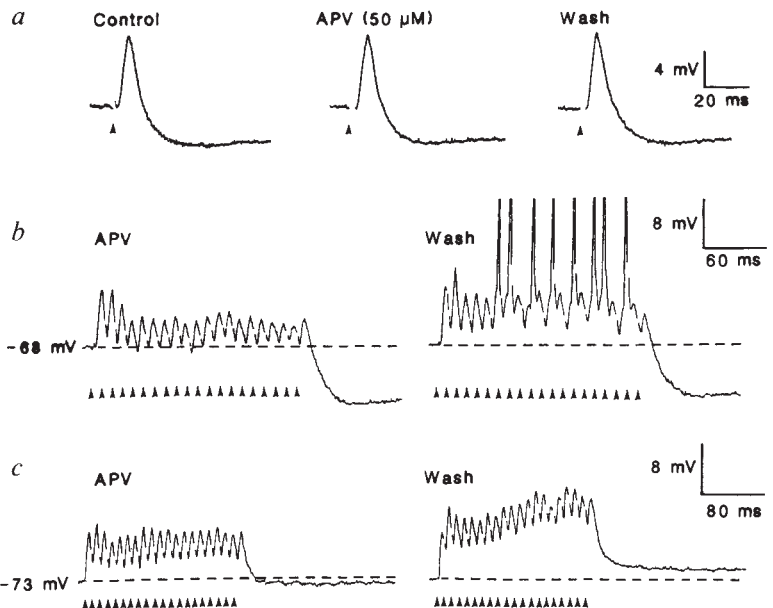


Fig. 2 Effect of 50 μ M APV on synaptic responses recorded intracellularly in CA1 neurones in response to stimulation of the Schaffer collateral-commissural pathway. *a*, Averages of five successive records of synaptic potentials evoked at 30-s intervals before (control), during and after washout of APV. *b*, Single records of responses of the same cell to high-frequency stimulation (20 shocks at 10-ms intervals) in the presence and following washout of APV. The e.p.s.p.s shown in *a* (centre and right-hand records) were obtained immediately before the two respective periods of high-frequency stimulation. *c*, Single records of responses to high-frequency stimulation in another cell. The times of stimulation are indicated by arrowheads and stimulus artefacts have been blanked out for clarity. Stimulus intensity and membrane potentials were constant throughout. Action potentials in *b* are truncated.

stimulation. It is not necessary, therefore, to postulate that high-frequency stimulation is required either to release an additional agonist for NMDA receptors or to increase neurotransmitter release sufficient to activate the NMDA receptors due to their localization or activation characteristics. A simpler hypothesis^{20,21} which agrees fully with the present observations is that a single neurotransmitter acts simultaneously on both non-NMDA and NMDA receptors but that the latter are prevented from participating appreciably in low-frequency synaptic transmission by Mg^{2+} . During high-frequency stimulation this Mg^{2+} blockade of the NMDA receptor system may be temporarily alleviated, allowing Ca^{2+} and Na^{+} to enter via the NMDA channels^{6,22-25}, and lead to the induction of LTP²⁶. This inward movement of cations through the NMDA channels could account for the slow e.p.s.p. recorded in the present study. It is likely that depolarization of the synaptic membrane, caused in part by temporal summation of fast e.p.s.p.s, could be the mechanism by which the blockade is reduced. In this respect, it has been shown using cultured neurones that Mg^{2+} directly blocks NMDA-gated channels and that this block is reduced by membrane depolarization²⁷. In addition, it has been reported recently that pairing single afferent synaptic inputs with post-synaptic depolarization can activate the NMDA receptor system and elicit LTP²⁸.

It has been shown that NMDA receptors are involved in epilepsy²⁹. In the presence of convulsant drugs, an NMDA receptor-mediated potential has been recorded extracellularly in the hippocampus during brief periods of high-frequency stimulation⁸, and a small NMDA receptor-mediated component of the fast e.p.s.p. recorded extracellularly²¹ or intracellularly^{13,14} has been detected during low-frequency stimulation. The latter effect has been explained on the basis of removal of the Mg^{2+} block of NMDA channels, the necessary depolarization being provided by the fast e.p.s.p.s prolonged by the action of the convulsant drug¹⁴. Such a mechanism could operate at the centre of an epileptic focus where synaptic inhibition may be impaired. The present results suggest a second way in which the NMDA receptor system may contribute to epileptic activity—by amplifying the synaptic output in response to a high-frequency input in the presence of fully functional synaptic inhibition (that is, absence of convulsant drugs). The synchronized, high-frequency discharge of neurones in epileptic tissue would provide ideal conditions for the activation of the NMDA receptor system such that this system could contribute to the propagation of epileptic activity through normal brain tissue.

It is now widely believed that NMDA receptors are important for synaptic plasticity in the hippocampus, but their function elsewhere in the brain is generally less clear. Although it is probable that NMDA receptors will be found to be involved in plastic processes in other regions of the CNS (for example, the cerebral cortex), their involvement in the discharge pattern of spinal interneurons¹, transmission of afferent sensory input³⁰ and generation of locomotor patterns³¹ is also indicated. As Mg^{2+} ions are a potent NMDA antagonist in all regions of the brain so far examined, it is likely that a synaptic mechanism similar to that proposed here for hippocampal neuronal transmission will apply throughout the CNS.

Since submission of this manuscript, it has been reported³² that postsynaptic hyperpolarization during conditioning reversibly prevents the initiation of LTP. These authors proposed that this was caused by the hyperpolarization preventing activation of voltage-gated calcium channels. On the basis of our present findings, we suggest a refinement of this hypothesis: that hyperpolarization prevented LTP by opposing the depolarization necessary to alleviate the Mg^{2+} block of NMDA channels.

This work was supported by the MRC and the Royal Society. We thank Dr J. C. Watkins for gifts of APV.

Received 6 January; accepted 28 April 1986.

1. Watkins, J. C. & Evans, R. H. A. *Rev. Pharmac. Tox.* **21**, 165-204 (1981).
2. Watkins, J. C. *Trends pharmac. Sci.* **5**, 373-376 (1984).
3. Olverman, H. J., Jones, A. W. & Watkins, J. C. *Nature* **307**, 460-462 (1984).
4. Crunelli, V., Forda, S., Collingridge, G. L. & Kelly, J. S. *Nature* **300**, 450-452 (1982).
5. Koerner, J. F. & Cotman, C. W. *Brain Res.* **251**, 105-115 (1982).
6. Collingridge, G. L., Kehl, S. J. & McLennan, H. J. *J. Physiol., Lond.* **334**, 33-46 (1983).
7. Collingridge, G. L., Kehl, S. J. & McLennan, H. J. *J. Physiol., Lond.* **338**, 27P (1983).
8. Wigström, H. & Gustafsson, B. *Neurosci. Lett.* **44**, 327-332 (1984).
9. Harris, E. W., Ganong, A. H. & Cotman, C. W. *Brain Res.* **323**, 132-137 (1984).
10. Slater, N. T., Stelzer, A. & Galvan, M. *Neurosci. Lett.* **60**, 25-31 (1985).
11. Morris, R. G. M., Anderson, E., Lynch, G. S. & Baudry, M. *Nature* **319**, 774-776 (1986).
12. Bliss, T. V. P. & Lomo, T. J. *J. Physiol., Lond.* **232**, 331-358 (1973).
13. Hynes, M. A. & Dingledine, R. *Soc. Neurosci. Abstr.* **10**, 229 (1984).
14. Herron, C. E., Williamson, R. & Collingridge, G. L. *Neurosci. Lett.* **61**, 255-260 (1985).
15. Collingridge, G. L., Kehl, S. J. & McLennan, H. J. *J. Physiol., Lond.* **334**, 19-31 (1983).
16. Coan, E. J. & Collingridge, G. L. *Neurosci. Lett.* **53**, 21-26 (1985).
17. Katz, B. & Miledi, R. *J. Physiol., Lond.* **168**, 389-422 (1963).
18. Hille, B., Woodhull, A. M. & Shapiro, B. I. *Phil. Trans. R. Soc. B270*, 301-318 (1975).
19. Hablitz, J. J. & Langmoen, I. A. *J. Neurosci.* **6**, 102-106 (1986).
20. Collingridge, G. L. *Trends pharmac. Sci.* **6**, 407-411 (1985).
21. Wigström, H., Gustafsson, B. & Huang, Y.-Y. *Acta physiol. scand.* **124**, 474-478 (1985).
22. Dingledine, R. *J. Physiol., Lond.* **343**, 385-405 (1983).
23. Heinemann, U., Hamon, B. & Konnerth, A. *Neurosci. Lett.* **47**, 295-300 (1984).
24. Ascher, P. & Nowak, L. *J. Physiol., Lond.* (in the press).
25. MacDermott, A. B., Mayer, M. L., Westbrook, G. L., Smith, S. J. & Barker, J. L. *Nature* **321**, 519-522 (1986).
26. Lynch, G., Larson, J., Kelso, S., Barrionuevo, G. & Schottler, F. *Nature* **305**, 719-721 (1983).

27. Nowak, L., Bregestovski, P., Ascher, P., Herbet, A. & Prochiantz, A. *Nature* **307**, 462-465 (1984).
28. Wigström, H., Gustafsson, B., Huang, Y.-Y. & Abrahams, W. C. *Acta physiol. scand.* **126**, 317-319 (1986).
29. Croucher, M. J., Collins, J. F. & Meldrum, B. S. *Science* **216**, 889-901 (1982).
30. Salt, T. E. *Nature* **322**, 263-265 (1986).
31. Dale, N. & Roberts, A. J. *Physiol., Lond.* **348**, 527-543 (1984).
32. Malinow, R. & Miller, J. P. *Nature* **320**, 529-530 (1986).

The inducible cytotoxic T-lymphocyte-associated gene transcript CTLA-1 sequence and gene localization to mouse chromosome 14

Jean-Francois Brunet, Magali Dosseto, François Denizot, Marie-Geneviève Mattei*, William R. Clark†, Tariq M. Haqqi, Pierre Ferrier, Markus Nabholz†, Anne-Marie Schmitt-Verhulst, Marie-Francoise Luciani & Pierre Golstein

Centre d'Immunologie INSERM-CNRS de Marseille-Luminy, Case 906, 13288 Marseille Cédex 9, France

* INSERM U.242, CHU Timone, 13385 Marseille Cedex 5, France.

† Swiss Institute for Experimental Cancer Research, Epalinges CH-1066, Switzerland.

Classical phenomenological approaches to the study of the mechanism of T-cell-mediated cytotoxicity¹⁻³ have now given way to a search for molecules involved in this function; this is attempted either by subcellular and biochemical fractionation of material from cytotoxic cells^{4,5}, or through the characterization of molecules recognized by cytotoxicity-inhibiting monoclonal antibodies (see ref. 6 for a review). Molecules having a role in cytotoxicity may also be identified by detecting the corresponding messenger RNA transcripts. Such an approach may include, as a first step, the search for transcripts as specific as possible to cytotoxic T cells; only secondarily can their actual relevance to cytotoxicity be investigated. We report here the preparation and systematic screening of a differential complementary DNA bank, in which we detected three distinct messenger RNA transcripts (CTLA-1, CTLA-2 and CTLA-3) present in various cytotoxic T cells but not (or less so) in a range of non-cytotoxic lymphoid cells. We describe the co-inducibility of these transcripts and of cytotoxicity in thymocytes and hybridoma cells, the sequence of CTLA-1 cDNA, its protein homology with serine esterases and the localization of the corresponding gene to mouse chromosome 14.

A library was derived from cDNA obtained by subtractive hybridization⁷⁻¹¹ between cDNA from the B10.BR anti-H-2K^b cytotoxic T-cell clone KB5C20 (ref. 12; hereafter designated T_c) and mRNA from the B-lymphoma M12.4.1 (ref. 13; hereafter designated B). The 8,300 recombinant cDNA clones obtained were subjected to three successive rounds of screening with radioactive cDNA probes prepared from a range of cytotoxic and non-cytotoxic cells. In the first round of screening, 230 of the 8,300 cDNA clones gave a signal with a T_c probe but not with a B probe. These 230 T_c⁺ B⁻ clones were subjected to a second round of screening using radioactive cDNA probes from T_c, B, T_h (T-helper cells, that is, the T-lymphoma EL4), thymocytes or brain. The results are summarized in Table 1. The 230 clones fell into a number of specificity patterns, with only 73 of them remaining 'T_c-specific'.

These 73 clones were hybridized with cDNA probes made from (1) RDM4, another non-cytotoxic T-cell lymphoma, detecting no clones, (2) EL4 cells, now activated with phorbol myristate acetate, which detected five 'activation-associated'

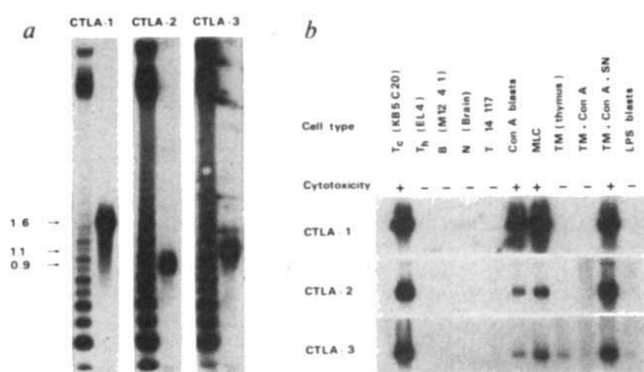


Fig. 1 Size and specificity patterns of CTLA-1, CTLA-2 and CTLA-3 RNA transcripts. Northern blots of RNA from the T_c clone KB5C20 (a) and from 11 cytotoxic and non-cytotoxic cell types (b) were probed with the inserts of plasmid M41D12, M41G12 or M11E6, representative of CTLA-1, CTLA-2 and CTLA-3, respectively. Cell types were as in Table 1, plus: T14.117, a hybridoma between a T-helper cell clone and the BW5147 thymoma²⁴; MLC cells, A/J spleen cells stimulated for 5 days *in vitro* with C57BL/6 irradiated spleen cells; TM + Con A, C57BL/6 thymocytes incubated for 24 h in the presence of 5 µg ml⁻¹ of Con A; TM + Con A + SN, TM + Con A further incubated for 5 days in the presence of IL-2-containing supernatant from PMA-stimulated EL4 C116 cells. At the time of RNA extraction, the cells were tested for cytotoxicity and classified as cytotoxic (+) or non-cytotoxic (-) according to the criteria given in Table 1. In a, an end-labelled 123-bp ladder (BRL) was used as a size marker. In b, a control actin probe gave a distinct band in all lanes (not shown). **Methods.** Total RNA (10 µg per lane) was run for 3-6 h at 2.5 V cm⁻¹ on a 1.1% formaldehyde-agarose gel, in MOPS buffer, then transferred to nitrocellulose filters. Inserts, obtained either from CsCl-purified plasmids or from minipreplications of plasmids, digested with *Pst*I and separated on 2% LMP (low melting point) agarose (BRL) gels, were labelled by random priming²⁵ and used as probes on Northern blots. Blots were prehybridized in 5×SSC, 50% formamide, 5×Denhardt's, 5 mM EDTA, 0.1% SDS, 50 mM phosphate buffer pH 7 and 7% dextran sulphate, at 42°C for 4-12 h. Hybridization was in the same buffer, to which the probe was added at 10⁶ d.p.m. ml⁻¹, for 12 h at 42°C. Washes were for 10 min in 2×SSC, 0.1% SDS and for 30 min in 0.2×SSC, 0.1% SDS at 65°C.

clones, (3) lipopolysaccharide (LPS)-stimulated blasts, detecting two clones, and (4) a mixture of 2-day-old and 3-day-old cytotoxic concanavalin A (Con A) blasts, which failed to detect nine cDNA clones that previously were found to be positive with a T_c probe. The latter clones might, therefore, be associated with long-term *in vitro* growth of T cells rather than with cytotoxicity. Also, 8 of the 73 T_c⁺ clones proved to be negative with a subtracted (T_c - B) cDNA probe, which was taken as an indication of non-T_c-specificity. Altogether, after this third round of screening, only 50 T_c-specific cDNA clones were retained for further study.

In order to identify clones corresponding to the same genes, inserts (of 200 to 1,400 base pairs, bp) from individual clones were purified and hybridized to all of the 50 clones. This approach identified five groups of cDNA clones. Of these, on the basis of transcript size as determined by Northern blotting (Fig. 1a), three groups were clearly distinct. These were designated CTLA-1, CTLA-2 and CTLA-3, and included at least 3, 3 and 39 cDNA clones, respectively. Transcripts were detected as single bands of 1.6 kilobases (kb) for CTLA-1, 0.9 kb for CTLA-2 and 1.1 kb for CTLA-3 (Fig. 1a). CTLA-1, CTLA-2 and CTLA-3 transcripts were present in all the cytotoxic cell populations tested (T-cell clones, mixed leukocyte culture cells, Con A blasts and thymocytes activated with Con A plus interleukin-2 (IL-2)-containing EL4 supernatant). The CTLA-1 transcript was not detected in brain cells or in a range of non-

† Permanent address: Molecular Biology Institute, University of California, Los Angeles, California 90024, USA.

Table 1 Pattern of specificity of selected clones from a subtracted (T_c - B) cDNA library

T_c	T_h	Probe TM	B	N	No. of cDNA clones
-	-	-	-	-	40
+	-	-	-	-	73
+	+	-	-	-	23
+	-	+	-	-	41
+	+	+	-	-	24
+	+	+	+	-	4
+	+	+	-	+	4
+	+	+	+	+	13

Two hundred and thirty $T_c^+B^-$ clones were transferred to filters and probed by hybridization with radioactive cDNA derived from, respectively, T_c (cytotoxic T-cell clone KB5C20), T_h (EL4 thymoma), TM (C57BL/6 thymocytes), B (M12.4.1 B-lymphoma cells) and N (C57BL/6 brain cells). + Indicates that a given cDNA clone hybridized detectably with the cDNA probe used, - that it did not. A few cDNA clones were lost during transfers and are not included. Forty clones did not give any positive signal and were considered false positives from the initial screening. Reciprocally, 17 clones gave a signal with the B probe and were considered false negatives from the initial screening. Only the 73 T_c -specific clones were subjected to further screening, although some of the rejected clones may be of interest in their own right.

Methods. Unselected total RNA was extracted by the LiCl method²⁰ from clone KB5C20 and the B lymphoma M12.4.1. Poly(A)⁺ RNA was selected by two passages over oligo(dT)-cellulose. KB5C20 single-stranded (SS) cDNA was synthesized from 5 μ g of poly(A)⁺ RNA in 100 mM Tris-HCl pH 8.3, 100 mM KCl, 10 mM MgCl₂, 10 mM dithiothreitol, 1 mM each of dA, dT and dG, 0.1 mM dC, 50 μ Ci ³²P-dCTP, oligo(dT)₁₀ at 2 U ml⁻¹ and 6 U of reverse transcriptase (super RT, Stehelin) for 1 h at 42 °C. Single-stranded cDNA (1.3 μ g) was hybridized to M12.4.1 poly(A)⁺ RNA (2 μ g μ l⁻¹) in 10 μ l of 0.4 M phosphate buffer pH 7.5, 1 mM EDTA, 0.1% SDS, overlaid with paraffin oil, for 20 h at 68 °C. After dilution in preheated 0.12 M phosphate buffer, the reaction was fractionated on a 1-ml hydroxyapatite column at 65 °C. The single-stranded fraction was precipitated in 50 mM EDTA pH 8, 0.1% cetylpyridinium bromide as described elsewhere²¹, except that an additional ethanol precipitation step was performed. The subtractive hybridization was repeated once, followed by an homologous hybridization with KB5C20 poly(A)⁺ RNA. The double-stranded fraction was precipitated as before, resuspended in H₂O and the RNA hydrolysed in 0.1 N NaOH. Thirty nanograms of single-stranded cDNA was recovered at this stage, that is, 6.4% of initial cDNA (notwithstanding nonspecific losses). Second-strand synthesis and elongation, S₁-nuclease digestion and dC tailing were performed as described elsewhere²². Fourteen nanograms of tailed cDNA were annealed to dGMP-tailed PstI-cut pBR322 plasmid (NEN) and used to transform *Escherichia coli* strain C600 according to the procedure of Hanahan²³. A total of 8,300 tetracycline-resistant, β -lactamase-negative (selected using Ampscreen, BRL) bacterial colonies containing recombinant plasmids were individually picked and grown into microtitre plate wells, replicated onto nitrocellulose filters and hybridized to ³²P-labelled cDNA synthesized as described above (however, with 50 μ M each of cold dG and dT, and 100 μ Ci each of ³²P-dT and ³²P-dG) from poly(A)⁺ RNA extracted from the following cells: T_c (KB5C20) and B (M12.4.1) for a first round of differential screening, and the same plus T_h (thymoma EL4 C116), TM (C57BL/6 thymus) and N (C57BL/6 adult brain) for a second round on the $T_c^+B^-$ clones selected in the first round. The results of this second round are presented above. Hybridizations were in 50% (v/v) formamide, 5 $\times$ SSC, 5 $\times$ Denhardt's, 200 μ g ml⁻¹ single-stranded DNA, for 48 h at 42 °C; washes were twice for 30 min each in 2 $\times$ SSC, 0.1% SDS and twice for 30 min each in 0.2 $\times$ SSC, 0.1% SDS at 65 °C. All the cells mentioned in this paper were tested, at the time of nucleic acid extraction, for their cytotoxic activity (or lack of it) in a conventional 4-h ⁵¹Cr-release test in the presence of the lectin Con A on both EL4 and RDM4 target cells. Cells were considered cytotoxic if they gave more than 20% specific ⁵¹Cr-release at a 20:1 effector to target cell ratio. In the absence of Con A, T_c specifically lysed EL4 but not RDM4 target cells, and induced PC60 cells lysed YAC target cells. To avoid contamination with feeder cell material, before extraction of RNA, T_c were grown for three passages without allogeneic feeder cells in the presence of EL4 C116 supernatant.

cytotoxic lymphoid cells (B-lymphoma cells, LPS blast cells, EL4 cells, normal thymocytes, Con-A-pulsed thymocytes, and hybridoma cells between a helper cell clone and BW5147 thymoma cells), even on overexposed films (Fig. 1b). The CTLA-2 transcript was detected in none of the non-cytotoxic cells at the exposure shown in Fig. 1b; with a longer exposure, faint bands appeared for all cells but the EL4 and B-lymphoma cells. The CTLA-3 transcript was detected in thymocytes and LPS blasts at the exposure shown in Fig. 1b, albeit less markedly than in the cytotoxic cells. These deviations from a strict T_c specificity had not been detected with the less sensitive methods used in the initial rounds of screening.

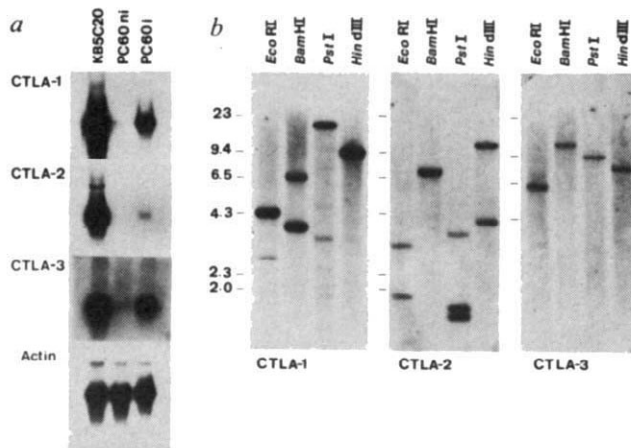


Fig. 2 Inducible transcription and Southern patterns of CTLA genes. *a*, Northern blots of RNA from KB5C20 (positive control) and non-induced (ni) or induced (i) PC60 hybridoma cells were probed with plasmid inserts representative of CTLA-1, CTLA-2 and CTLA-3, respectively, and with a control actin probe (given by M. Buckingham via C. Goridis); induction of PC60 cells also led to the appearance of cytotoxicity (not shown). Exposure times were different for each probe. *b*, Southern blots of DNA from mouse liver cells, cut with the indicated restriction enzymes, were probed with the same plasmid inserts; the numbers on the left are *Hind*III-derived size markers (in kb). Identical patterns were obtained with BALB/c liver, B10.BR liver or KB5C20 DNA (not shown).

Methods. Northern blots were performed and probed as described in Fig. 1 legend. PC60 hybridoma cells were induced by adding IL-1 and IL-2-containing supernatant to PC60 cells as described previously²⁶ (we used the PC60-21-14-4 subclone of PC60). For Southern blots, DNA was prepared and purified by centrifugation to equilibrium in CsCl essentially as described elsewhere²⁷, except that a higher concentration of Tris and EDTA was used in homogenizing buffer. Genomic DNA (10 μ g) was digested to completion with *Eco*RI, *Bam*HI, *Pst*I or *Hind*III according to the conditions specified by the manufacturer (Boehringer Mannheim), electrophoresed for 16 h in 0.8% agarose gel in TAE (tris-acetate-EDTA) buffer and transferred to Gene Screen Plus (Du Pont). Blots were prehybridized for 6 h at 65 °C in 6 $\times$ SSC, 5 $\times$ Denhardt's, 1% SDS, 10 mM EDTA and 150 μ g ml⁻¹ denatured salmon sperm DNA. Probes were labelled²⁵ to high specific activity (1.6 $\times$ 10⁹ c.p.m. μ g⁻¹). Hybridization medium was as above plus 10% dextran sulphate and denatured probe (8 $\times$ 10⁶ c.p.m. ml⁻¹). Blots were hybridized overnight, then washed down to 0.2 $\times$ SSC at 65 °C for 100 min, dried and exposed to film overnight.

Expression of the CTLA transcripts was induced under conditions (addition of IL-2) which led to the differentiation of non-cytotoxic thymocytes into cytotoxic cells (Fig. 1b); this, however, might reflect the differential growth on stimulation of a small subpopulation of precursor cells already making these transcripts. We therefore used PC60 hybridoma cells¹⁴ which are constitutive for growth but inducible for cytotoxicity; all three CTLA transcripts, which in non-induced PC60 cells were not or only barely detectable, appeared in PC60 cells when cytotoxicity was induced (Fig. 2a). Thus, in both thymocytes and hybridoma cells there was a correlation between induction of cytotoxicity and induction of the three CTLA transcripts.

We then wondered whether CTLA-1, which gave the clearest association (Fig. 1b) and induction (Fig. 2a) results, was homologous or identical to a previously sequenced molecule. The cDNA sequence that we obtained (Fig. 3) was not significantly homologous to any of the sequences in the Los Alamos Gene Bank. However, the translated protein sequence showed significant homology, but not identity, with known serine esterases (Fig. 3).



Fig. 3 The sequence of CTLA-1 cDNA and its protein translation. **a**, Sequencing strategy. The open reading frame of 741 nucleotides is indicated by a box on clone 7.1.2. **b**, cDNA sequence and the predicted amino-acid sequence. The open reading frame starts from the first ATG downstream of the stop codon at position 79. The 3'-untranslated region ends with the consensus polyadenylation signal AATAAA. There is a potential *N*-linked glycosylation site at Asn 70. The conceptual amino-acid sequence shows clustered homologies with trypsin-related serine-esterases, in particular around the serine-esterase active-site residues His, Asp and Ser (here at positions 64, 108 and 203, respectively). **c**, Alignment of the rat group-specific serine esterase sequence²⁸ with that of CTLA-1 from Ile 21 to Ile 240. The overall homology is 46%; identical stretches of sequence are boxed. Given the highly hydrophobic leader sequence in the NH₂-terminal region, the potential site of cleavage of an activation peptide at Arg 15 or perhaps rather Lys 17, and the sequence homology with the initial residues of serine esterases, the putative active CTLA-1 protein would start at Ala 18 and therefore be 230 amino acids long.

Methods. KB5C20-derived single-stranded cDNA was synthesized as described in Table 1 (except that 4 mM Na-PP_i was included in the reaction), made double-stranded using RNase H (BRL), Pol I (Boehringer) and *E. coli* lipase (BRL) as described by Gubler and Hoffman²⁹ (except that the incubation was for 1 h at 12 °C and 1 h at 16 °C), tailed by TdT (NEN; 20 U µg⁻¹ in 50 µl of 200 mM K₂CO₃ pH 7.2, 1 mM CoCl₂, 500 µg ml⁻¹ bovine serum albumin, 5 µM dCTP for 40 min at 15 °C), and size-fractionated on 1.5% agarose. Slices corresponding to 1.2–1.6 kb were electroeluted and double-stranded cDNA annealed to the *Pst*I fragment of dG-tailed PUC9 (NEN); transformation was as described in Table 1. The library was screened with the 450-bp *Pst*I fragment of M41D12 which detected 10 clones, of which the 3 longest were analysed. Sequences were obtained by the dideoxy chain-termination method³⁰ for the two *Pst*I fragments of M41D12 subcloned in M13, and by chemical cleavage³¹ following the strategy shown in **a**.

Southern patterns of liver DNA tested with probes from the three CTLA families showed a small number of bands (Fig. 2b), indicating that the corresponding genes are present as either one copy or a small number of copies per haploid genome. The respective patterns were the same for KB5C20 DNA (not shown), thus providing no detectable evidence for CTLA gene rearrangements in T cells. The chromosomal localization of the CTLA-1 gene was determined by *in situ* hybridization. Only one major binding site was found, in the D segment of mouse chromosome 14 (Fig. 4), interestingly in a region to which the T-cell receptor (TCR) α -chain gene has been mapped¹⁵. Experiments in progress, taking advantage of the cross-hybridization of CTLA-1 mouse probes to human DNA (not shown), suggest that CTLA-1 and TCR α genes might be co-localized on human chromosomes also.

Thus, we have detected three 'CTLA' genes whose transcription is induced preferentially or exclusively in cytotoxic T lymphocytes. A first question raised by these results is the nature of the CTLA protein products. That CTLA-1 is a cytotoxic T-cell-associated serine esterase is reminiscent of the detection

in cytotoxic T cells of such enzymes at the protein level¹⁶. A comparison of partial sequences showed that CTLA-1 and the C11 clone described by Lobe *et al.*¹⁷ were almost identical; differences in the Northern patterns are being investigated. A second, key question—whether the CTLA products are actually involved in the mechanism of cytotoxicity—could be investigated by using anti-sense mRNA technology^{18,19}; the probable serine esterase activity of CTLA-1 suggests that it might have a profactor-activating role. Irrespective of the answers to these questions, it will be interesting to study the regulation of expression of the CTLA-1 gene in particular, which maps near the TCR α gene and is induced on differentiation into cytotoxic T-cells.

We thank C. Goridis, B. Jordan, B. Malissen and M. Pierres for advice and for reading of the manuscript, M. Pierres for the gift of T14.117 hybridoma cells, J. Sire and R. Sodoyer for advice and help in DNA sequencing, and M. Masson (Pasteur Institute) for computer searches in gene and protein sequence banks. We thank Dr Bleackley (Edmonton) for a crucial suggestion and for a comparison of a partial sequence. W. R. C. was supported by an Eleanor Roosevelt Fellowship and T.M.H. by the Founda-

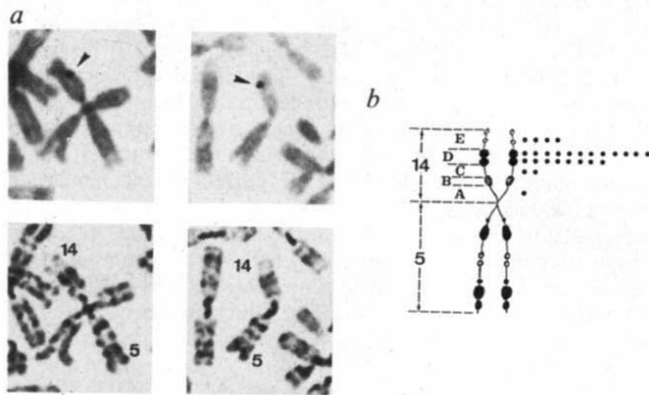


Fig. 4 Localization of the CTLA-1 gene to mouse chromosome 14 by *in situ* hybridization with the CTLA-1 probe C.1.1. *a*, Two partial WMP mouse metaphases, showing the specific site of hybridization to chromosome 14. Top, arrowheads indicate silver grains on Giemsa-stained chromosomes, after autoradiography. Bottom, chromosomes with silver grains were subsequently identified by R-banding. *b*, Diagram of WMP mouse Rb(5; 14) chromosome, indicating the distribution of labelled sites. In the 85 metaphases examined, there were 152 silver grains associated with chromosomes and 27 of these (17.7%) were localized on chromosome 14. The distribution of grains on this chromosome was not random: 74% of them mapped to the D band, with a maximum in the distal D3 part of this band (for banding nomenclature, see ref. 32).

Methods. *In situ* hybridizations with probe C.1.1 were done using metaphase spreads from a male WMP mouse in which all the autosomes (except 19) are in the form of metacentric robertsonian translocations (given by J. L. Guénet, Pasteur Institute, Paris). Con A-stimulated lymphocytes were cultured at 37 °C for 72 h, with 5-bromodeoxyuridine added for the final 7 h of culture (30 µg per ml of medium) to ensure a chromosomal R-banding of good quality. Total C.1.1. plasmid DNA was tritium-labelled by nick-translation to a specific activity of 10^8 c.p.m. µg⁻¹. The radiolabelled probe was hybridized to metaphase spreads at a final concentration of 2 µg per ml of hybridization solution, as described previously³³. After coating with nuclear track emulsion (Kodak NTB 2), the slides were exposed for 8 days at 4 °C, then developed. To avoid any slipping of silver grains during the banding procedure, chromosome spreads were first stained with buffered Giemsa solution and metaphases photographed. Then chromosome banding was realized by the fluorochrome-photolysis-Giemsa (FPG) method described by Camargo and Cervenka³⁴, and metaphases photographed again before analysis.

tion pour la Recherche Médicale. Other support was through institutional grants from INSERM and CNRS and through a specific grant from the Association pour la Recherche contre le Cancer.

Received 7 April; accepted 25 April 1986.

- Henney, C. S. *Contemp. Topics Immunobiol.* **7**, 245-272 (1977).
- Golstein, P. & Smith, E. T. *Contemp. Topics Immunobiol.* **7**, 273-300 (1977).
- Mart, E. *Contemp. Topics Immunobiol.* **7**, 301-361 (1977).
- Henkart, P. A., Millard, P. J., Reynolds, C. W. & Henkart, M. P. *J. exp. Med.* **160**, 75-93 (1984).
- Podack, E. R. & Konigsberg, P. J. *J. exp. Med.* **160**, 695-710 (1984).
- Golstein, P. *et al. Immun. Rev.* **68**, 5-42 (1982).
- Timberlake, W. E. *Dev. Biol.* **78**, 497-510 (1980).
- Yanagi, Y. *et al. Nature* **308**, 145-149 (1984).
- Hedrick, S. M., Cohen, D. I., Nielsen, E. A. & Davis, M. *Nature* **308**, 149-153 (1984).
- Saito, H. *et al. Nature* **309**, 757-762 (1984).
- Sims, J. E., Tunncliffe, A., Smith, W. J. & Rabbitts, T. H. *Nature* **312**, 541-545 (1984).
- Albert, F., Buferne, M., Boyer, C. & Schmitt-Verhulst, A.-M. *Immunogenetics* **16**, 533-549 (1982).
- Hamano, T., Kim, K. J., Leiserson, W. M. & Asofsky, R. *J. Immun.* **129**, 1403-1406 (1982).
- Conzelmann, A., Corthésy, P., Gianfriglia, M., Silva, A. & Nabholz, M. *Nature* **298**, 170-172 (1982).
- Kranz, D. M. *et al. Science* **227**, 941-945 (1985).
- Pasternack, M. S. & Eisen, H. N. *Nature* **314**, 743-745 (1985).
- Lobe, C. G., Havele, C. & Bleackly, R. C. *Proc. natn. Acad. Sci. U.S.A.* **83**, 1448-1452 (1986).
- Izant, J. G. & Weintraub, H. *Cell* **36**, 1007-1015 (1984).
- Rubinstein, J. L. R., Nicolas, J.-F. & Jacob, F. *C. r. heb. Acad. Séanc. Acad. Sci., Paris* **299**, 271-274 (1984).

- Auffray, C. & Rougeon, R. *Eur. J. Biochem.* **107**, 303-314 (1980).
- Geck, P. & Nasz, I. *Analyt. Biochem.* **135**, 264-268 (1983).
- Maniatis, T., Fritsch, E. F. & Sambrook, J. in *Molecular Cloning. A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
- Hanahan, D. *J. molec. Biol.* **166**, 557-580 (1983).
- Pont, S. *et al. Eur. J. Immun.* **15**, 1222-1228 (1985).
- Feinberg, A. & Vogelstein, B. *Analyt. Biochem.* **137**, 266-267 (1984).
- Erard, F. *et al. J. exp. Med.* **160**, 584-599 (1984).
- Kaiser, K. & Murray, N. E. in *DNA Cloning* Vol. 1 (ed. Glover, D. M.) (IRL Press, Oxford, 1985).
- Woodbury, G., Katunuma, N., Kobayashi, K., Titani, K. & Neurath, H. *Biochemistry* **17**, 811-819 (1978).
- Gubler, U. & Hoffman, B. J. *Gene* **25**, 263-269 (1983).
- Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
- Chuvpilo, S. A. & Kravchenko, V. V. *FEBS Lett.* **179**, 34-36 (1984).
- Nesbitt, M. N. & Francke, U. *Chromosoma* **41**, 145-158 (1973).
- Mattéi, M. G. *et al. Hum. Genet.* **69**, 268-271 (1985).
- Camargo, M. & Cervenka, J. *Am. J. hum. Genet.* **34**, 757-780 (1982).

Voltage-gated Ca²⁺ entry into *Paramecium* linked to intraciliary increase in cyclic GMP

Joachim E. Schultz, Thomas Pohl & Susanne Klumpp

Pharmazeutisches Institut der Universität Tübingen, Morgenstelle 8, 7400 Tübingen, FRG

Inward Ca²⁺ currents exist in many excitable tissues and are linked to regulation of several cellular processes such as cyclic GMP formation¹⁻⁴. In *Paramecium*, a graded Ca²⁺/K⁺ action potential regulates swimming behaviour⁵⁻⁷. Voltage-gated Ca²⁺ channels localized in the excitable ciliary membrane^{8,9} conduct a depolarizing influx of Ca²⁺ and translate changes in membrane potential into a transient Ca²⁺ signal which triggers ciliary reversal, that is, backward swimming^{5,6}. A guanylate cyclase that is activated specifically by Ca²⁺ has already been characterized in the ciliary membrane¹⁰. By using behavioural mutants of *Paramecium* with reduced^{6,11,12} or exaggerated^{13,14} Ca²⁺ currents, we now demonstrate in an intact animal a direct link between the voltage-gated inward Ca²⁺ current and an elevation of cyclic GMP levels. Although the increased cyclic GMP level does not directly influence swimming behaviour, the combination of electrophysiology, biochemistry and genetics possible with *Paramecium* offers an opportunity of identifying the role of cyclic GMP levels in cell behaviour.

In several *Paramecium* strains, increasing the external K⁺ concentration from 5 to 25 mM enhances ciliary Ca²⁺ permeability and causes a reversal of ciliary beating and consequently of the swimming direction⁵⁻⁷, a response similar to the avoidance displayed by *Paramecium* in adverse situations. This reversal lasts for 5-10 s and is mediated by the influx of Ca²⁺ into the cilia. In *Paramecium tetraurelia* wild-type 51s, the concentration of cyclic GMP increases from 3.8 to 5.5 pmol per mg protein 3-5 s after the addition of K⁺ (Fig. 1a). The increase is moderate because the influx of Ca²⁺ leads to an inactivation of the Ca²⁺ channel within 5 ms^{5,15}. Consequently, the Ca²⁺-dependent guanylate cyclase is only briefly stimulated. Therefore, Ba²⁺ was used to increase the excitability of the Ca²⁺ channels by decreasing their rate of inactivation, thereby producing a more sustained (if somewhat slower) Ca²⁺ entry^{11,15,16}. Addition of 3 mM Ba²⁺ to wild-type *Paramecium* evoked a rapid, large and transient increase in cyclic GMP (Fig. 1b). The level of cyclic GMP attained a maximum 5-10 s after stimulation and returned towards resting levels after 30 s. In the presence of Ba²⁺ the backward swimming response continued for ~60 s, indicating that the level of cyclic GMP does not directly control backward swimming. The effect of Ba²⁺ was specific, as 3 mM of Mg²⁺, Ca²⁺ or Sr²⁺ did not elicit the avoidance reaction or an increase in cyclic GMP. Moreover, the temperature dependence of the Ba²⁺-stimulated increase in cyclic GMP indicates that it is not the ion current itself but a subsequent step that is responsible for the effect. At 0 °C Ba²⁺ does not evoke an increase in cyclic

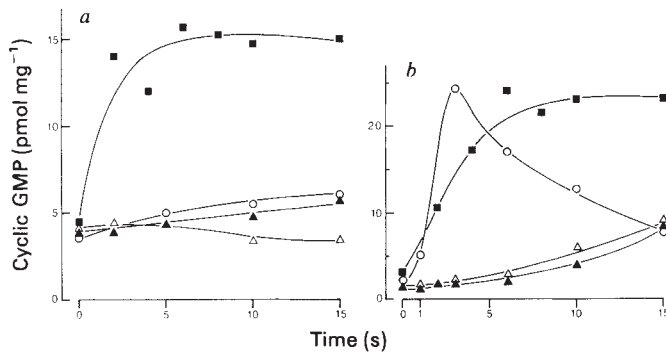


Fig. 1 Time-dependent increase in cyclic GMP concentrations in *Paramecium tetraurelia* wild-type 51 s (○), mutant pawn A (d4-94, △), double mutant pawn A/pawn B (▲) and mutant dancer (d4-614, ■) after a sudden increase in the concentration of K^+ from 5 to 25 mM (a) or after addition of 3 mM Ba^{2+} (b).

Methods. Early stationary cells grown in a bacterized hay infusion or in axenic complex medium²⁶ were collected by centrifugation at 250g, washed and equilibrated for 2 h at 40,000 cells ml^{-1} (90,000 cells = 1 mg protein) in buffer containing 50 μM Ca^{2+} , 5 mM K^+ , 10 mM MOPS pH 7.2. For stimulation, 250 μl of 40 mM K^+ (a) or 6 mM $BaCl_2$ (b) in equilibration buffer were added to 250 μl of cells. Viability and behaviour were checked microscopically before and during the experiments. Incubations were stopped by adding perchloric acid (final concentration 1 M). Cyclic GMP was determined by radioimmunoassay using rabbit antibodies against cyclic GMP and ^{125}I -iodinated 2'-O'-succinyl-cyclic GMP-L-tyrosinylmethyl ester (for details see ref. 27). Cross-reactivity of the antibody with cyclic AMP was <5%. No cyclic GMP was found when samples were digested with phosphodiesterase isolated from pig brain before the assay. We did not detect cyclic GMP in the incubation medium nor did we observe differences in behaviour or cyclic GMP levels between cells grown in bacterized medium and those grown axenically.

GMP, although influx of Ca^{2+} does occur¹⁷; we attribute this effect to the lower activity of guanylate cyclase at this temperature.

Because both Ba^{2+} and Ca^{2+} compete for the inactivation site on the channel, the electrophysiological and behavioural effects of Ba^{2+} stimulation depend on the ratio of these cations in the medium^{16,18}. We established dose-response curves for Ba^{2+} in the presence of various Ca^{2+} concentrations. The ED_{50} (50% effective dose) values for Ba^{2+} -stimulated cyclic GMP levels in the presence of 3, 50 and 1,000 μM Ca^{2+} were 0.1, 2 and 10 mM, respectively (Fig. 2), reflecting the competition between Ca^{2+} and Ba^{2+} . Ba^{2+} elicited a concentration of cyclic GMP of only 15 $pmol\ mg^{-1}$ at 3 μM Ca^{2+} , compared with 28 $pmol\ mg^{-1}$ at an external Ca^{2+} concentration greater than 50 μM (Fig. 2). Thus, low external Ca^{2+} , leading to reduced Ca^{2+} influx^{16,19}, is correlated with considerably reduced production of intracellular cyclic GMP. Obviously, release of Ca^{2+} from intraciliary storage sites, if it occurred, was not involved in stimulating cyclic GMP production. Ba^{2+} itself does not stimulate guanylate cyclase activity *in vitro*²⁰ or antagonize Ca^{2+} -mediated activation (Fig. 3). Importantly, the ED_{50} of 9 μM Ca^{2+} for guanylate cyclase activation was not affected by the presence of 1 mM Ba^{2+} (Fig. 3). Thus we conclude that during an action potential Ca^{2+} enters the cilia with Ba^{2+} and directly stimulates ciliary guanylate cyclase.

Further evidence for a link between the inward Ca^{2+} current and intraciliary cyclic GMP production can be obtained by studying mutants possessing defects in ion conductances^{5-7,11,12}. 'Pawn' mutants lack voltage-gated Ca^{2+} channels^{11,12,16} and therefore cannot swim backwards. In agreement with electrophysiological and behavioural observations^{6,11,12}, we found that addition of 3 mM Ba^{2+} to pawn A or pawn A/pawn B cells did not stimulate cyclic GMP biosynthesis as in wild-type cells (Fig. 1b). 'Atalanta' mutants lack the axonemal response to Ca^{2+}

despite having normal Ca^{2+} channels, so they are unable to reverse the direction of the ciliary beat and cannot swim backwards^{12,21}. Addition of 3 mM Ba^{2+} to atalanta cells produced an increase in the cyclic GMP concentration, from 4.6 $pmol\ mg^{-1}$ to 18.0 $pmol\ mg^{-1}$ after 10 s; as expected, this is comparable to the response of wild-type cells and reflects the fact that atalanta mutants have an intact inward Ca^{2+} current. Dancer mutants possess Ca^{2+} channels which inactivate more slowly and incompletely, resulting in enhanced inward Ca^{2+} currents^{13,14}. 'Dancer' cells display much longer periods of backward swimming when depolarized by K^+ than wild-type *Paramecium*¹³. Thus, K^+ stimulation should have a profound effect on the cyclic GMP content of dancer cells if the inward Ca^{2+} current is linked to intraciliary guanylate cyclase activation. Indeed, K^+ evoked a fourfold increase in the concentration of cyclic GMP in dancer cells (Fig. 1a). In Ba^{2+} -stimulated dancer cells, levels of cyclic GMP remained higher than those in comparably stimulated wild-type *Paramecium* (Fig. 1b). These observations are compatible with the defect in Ca^{2+} -channel inactivation in the dancer cells¹³.

The differences in cyclic GMP production *in vivo* between wild-type and mutant *Paramecium* could be caused by different activities of guanylate cyclase. Therefore, enzyme activities were measured in cilia from all *Paramecium* strains used. The mean guanylate cyclase activity in wild-type and mutant preparations was 175 $pmol$ per min per mg. We also investigated a possible involvement of phosphodiesterase in the regulation of cyclic GMP concentration. Using partially purified cyclic GMP phosphodiesterase from *Paramecium*, we found no effect of Ca^{2+} or Ba^{2+} . Thus, inward flow of Ca^{2+} apparently stimulates intraciliary cyclic GMP production directly. Previously, we have shown by immunocytochemistry that cyclic GMP is predominantly localized in the cilia of *Paramecium*²²; this corroborates the link between voltage-gated Ca^{2+} influx and Ca^{2+} -regulated guanylate cyclase.

Sophisticated control loops involving Ca^{2+} and cyclic nucleotides probably operate in most cells^{23,24}. In particular, hormonally stimulated production of cyclic GMP in metazoan tissues—for example, in brain¹, smooth muscle² and neuroblastoma cells³—has been demonstrated to depend on the presence of Ca^{2+} . However, guanylate cyclases regulated by physiologically meaningful Ca^{2+} concentrations have not been found.

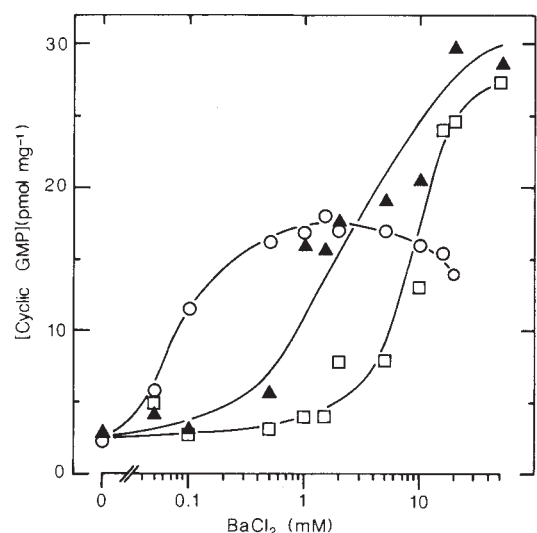


Fig. 2 Dose-response curves for Ba^{2+} -stimulated cyclic GMP production in *Paramecium tetraurelia* wild-type 51 s in the presence of 3 (○, $n=5$), 50 (▲, $n=13$) and 1,000 (□, $n=3$) μM Ca^{2+} in the equilibration and incubation buffer. Additions of Ba^{2+} were made to yield the indicated concentrations without affecting the concentration of other buffer constituents. Measurements were made 10 s after Ba^{2+} addition. For methods see Fig. 1 legend.

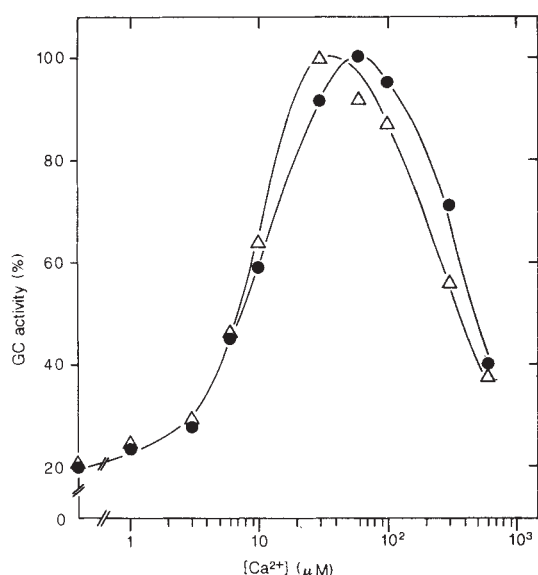


Fig. 3 Dose-response curve for Ca^{2+} -stimulated guanylate cyclase (GC) in ciliary membranes from *Paramecium*. Membranes were washed with $60\ \mu\text{M}$ EGTA followed by two washing steps with buffer to remove the chelator. All reagents were purified by treatment with Chelex. The remaining Ca^{2+} concentration was $<1\ \mu\text{M}$. Activity was determined for various concentrations of Ca^{2+} in the absence (●) or presence (Δ) of $1\ \text{mM}$ Ba^{2+} . 100% activity corresponds to $1,560\ \text{pmol}$ cyclic GMP per mg per min.

Here, we have demonstrated for the first time an unequivocal link between an inward Ca^{2+} current and enhanced cyclic GMP production.

In *Paramecium* the voltage-gated Ca^{2+} channels in the ciliary membrane serve as a mediator of all excitatory (that is, depolarizing) stimuli. In response to depolarizing receptor potentials, an action potential starts with a Ca^{2+} influx. Internally, Ca^{2+} has at least four effects: the inactivation of the Ca^{2+} channels within 5 ms, the activation of a Ca^{2+} -dependent K^{+} efflux after 100 ms (ref. 25), the reversal of the direction of the ciliary power stroke, and the production of the second messenger cyclic GMP. That the electrophysiology, genetics and biochemistry of *Paramecium* can be studied in combination holds great promise for elucidating the molecular events of basic cellular processes controlled by cyclic GMP. At present, there are insufficient data to assign a common function for cyclic GMP in all cells but the functional relationship between cyclic GMP and Ca^{2+} in many tissues (for example in rod outer segments, which are phylogenetically a rudimentary cilium⁴) suggests that similarities in the regulation of cyclic GMP in different cellular systems will be found.

We thank Dr C. Kung (University of Wisconsin, Madison) for the *Paramecium* mutants. This work was supported by the Deutsche Forschungsgemeinschaft (SFB 323).

Received 19 February; accepted 21 April 1986.

1. Ferrendelli, J. A., Rubin, E. H. & Kinscherf, D. A. *J. Neurochem.* **26**, 741-748 (1976).
2. Takayanagi, I., Hisayama, T. & Kotsugai, T. *Jap. J. Pharmac.* **31**, 831-834 (1981).
3. El-Fakahany, E. & Richelson, E. *Proc. natn. Acad. Sci. U.S.A.* **77**, 6897-6901 (1980).
4. Cohen, A. I. *Curr. Topics Membranes Transp.* **15**, 215-229 (1981).
5. Eckert, R. & Brehm, P. A. *Rev. Biophys. Bioeng.* **8**, 353-383 (1979).
6. Kung, C. & Saimi, Y. A. *Rev. Physiol.* **44**, 519-534 (1982).
7. Naitoh, Y. in *Electrical Conduction and Behaviour in 'Simple' Invertebrates* (ed. Shelton, A. B.) 1-48 (Clarendon, Oxford, 1982).
8. Ogura, A. & Takahashi, K. *Nature* **264**, 170-172 (1976).
9. Dunlap, K. J. *Physiol., Lond.* **271**, 119-133 (1977).
10. Klumpp, S. & Schultz, J. E. *Eur. J. Biochem.* **124**, 317-324 (1982).
11. Oertel, D., Schein, S. J. & Kung, C. *Nature* **268**, 120-124 (1977).
12. Kung, C., Chang, S. Y., Satow, Y., van Houten, J. & Hansma, H. *Science* **188**, 898-904 (1975).
13. Hinrichsen, R. D. & Saimi, Y. *J. Physiol., Lond.* **351**, 397-410 (1984).
14. Hinrichsen, R. D., Saimi, Y. & Kung, C. *Genetics* **108**, 545-558 (1984).
15. Brehm, P., Eckert, R. & Tillotson, D. J. *Physiol., Lond.* **306**, 193-203 (1980).

16. Ling, K.-Y. & Kung, C. *J. exp. Biol.* **84**, 73-87 (1980).
17. Browning, J. L., Nelson, D. L. & Hansma, H. *Nature* **259**, 491-494 (1976).
18. Naitoh, Y. & Eckert, R. Z. *vergl. Physiol.* **8**, 453-472 (1968).
19. Satow, Y. & Kung, C. *J. exp. Biol.* **78**, 149-161 (1979).
20. Klumpp, S., Kleefeld, G. & Schultz, J. E. *J. biol. Chem.* **258**, 12455-12459 (1983).
21. Hinrichsen, R. D. & Kung, C. *Genet. Res.* **43**, 11-20 (1984).
22. Klumpp, S., Steiner, A. L. & Schultz, J. E. *Eur. J. Cell Biol.* **32**, 164-170 (1983).
23. Rapp, P. E. & Berridge, M. J. *J. theor. Biol.* **66**, 497-525 (1977).
24. Berridge, M. J. *Adv. Cyclic Nucleotide Prot. Phosphor.* **17**, 329-335 (1984).
25. Satow, Y. & Kung, C. *J. exp. Biol.* **88**, 293-303 (1980).
26. Van Wagtenonk, W. J. in *Paramecium: A Current Survey* (ed. van Wagtenonk, W. J.) 339-378 (Elsevier, Amsterdam, 1974).
27. Delaage, M. A., Roux, D. & Cailla, H. L. *Molecular Biology and Pharmacology of Cyclic Nucleotides* (eds Falco, G. & Paoletti, R.) 155-170 (Elsevier, Amsterdam, 1978).

Expression of peptide chain release factor 2 requires high-efficiency frameshift

William J. Craigen & C. Thomas Caskey

Institute for Molecular Genetics, Department of Biochemistry, and Howard Hughes Medical Institute, Baylor College of Medicine, Houston, Texas 77030, USA

Peptide chain release factors are soluble proteins that participate in the stop codon-dependent termination of polypeptide biosynthesis. In *Escherichia coli*, two release factors are necessary for peptide chain termination: release factor 1 (RF1) specifies UAG- and UAA-dependent termination whereas release factor 2 (RF2) specifies UGA- and UAA-dependent termination¹. Release factors are found in low concentrations relative to other translation factors², suggesting that their expression is tightly regulated and, accordingly, making the study of their structure-function relationship difficult. RF1 and RF2 exhibit significant sequence homology, probably reflecting their similar functions and perhaps a common evolutionary origin³. DNA and peptide sequencing have suggested the existence of a unique mechanism for the autogenous regulation of RF2 in which an in-frame UGA stop codon requires an obligatory +1 frameshift within the coding region of the RF2 gene. In this report we present *in vitro* experimental results consistent with the autogenous regulation of RF2. Additionally, we used RF2-lacZ gene fusions to demonstrate that autogenous regulation occurs, at least in part, by premature termination at the in-frame stop codon, since deletion of this stop codon leads to overproduction of the RF2-LacZ fusion protein. Frameshifting at this premature termination codon occurs at the remarkably high rate of 50%.

DNA sequencing demonstrated that the RF2 gene contains a UGA (*opal*) stop codon in-frame with the initiator methionine codon at codon position 26. Peptide sequencing of the amino terminus of RF2 confirmed that a +1 frameshift must occur at this position for the complete translation of intact RF2³. Since RF2 uniquely recognizes *opal* stop codons⁴, RF2 may autoregulate when present in excess by prematurely terminating its own synthesis at codon 26.

To test this hypothesis we have used an *in vitro* coupled transcription/translation system to determine whether excess RF2 protein can suppress the translation of plasmid-encoded RF2. This system has been used to demonstrate autogenous translational regulation for a number of ribosomal protein operons⁵. Figure 1 shows the result of adding varying amounts of purified RF2 to the reaction mixture. Using the plasmid-encoded β -lactamase as an internal standard it can be seen that RF2 production is reduced by increasing concentrations of added RF2. The addition of purified RF1 has no measurable effect (results are summarized in Fig. 2), and conversely the addition of RF2 has no measurable effect on RF1 synthesis (data not shown). We conclude that, at least *in vitro*, RF2 is regulated at the translational level by its own gene product.

To determine whether this in-frame stop codon is the actual regulatory mechanism, and to measure the rate of frameshifting

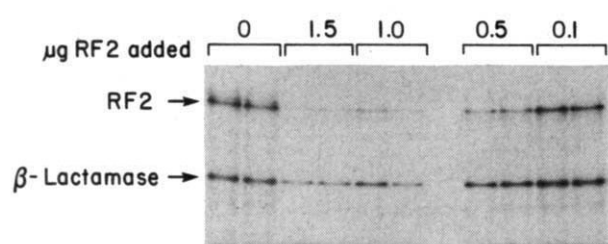


Fig. 1 *In vitro* autogenous regulation of RF2. A coupled transcription/translation system (Amersham) is used to demonstrate autogenous regulation of RF2. To a 30- μ l reaction mixture containing 35 S-methionine, the indicated amount of purified RF2 protein¹⁵ was added and the reaction started by the addition of 1 μ g pRF2-1. The reaction was incubated at 37 °C for 30 min, followed by a 15-min chase with cold methionine, and the reaction products were displayed on a 12% SDS-polyacrylamide gel. The overall efficiency of the reaction is somewhat inhibited by the addition of RF2. Each reaction was performed in duplicate, and the complete experiment performed three times.

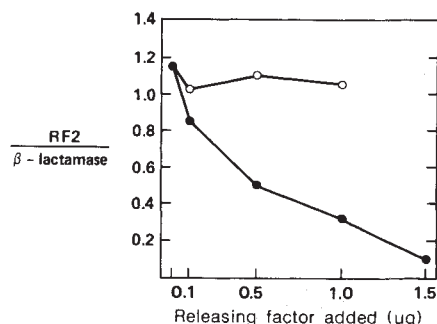


Fig. 2 A summary of the *in vitro* translation data for RF1 (○) and RF2 (●). Band intensity was measured with a Helena Laboratories Auto-scanner densitometer and the intensity plotted as a ratio of RF2 to β -lactamase. The effect of purified RF1 protein was assessed as described in Fig. 1 legend for RF2, and plotted similarly. A representative experiment is depicted in the graph.

at this stop codon, *RF2-lacZ* gene fusions were constructed both with and without codon 26. It was assumed that these fusion proteins would be inactive for release factor activities and thus would not contribute to their own regulation. The construction of the gene fusions is shown diagrammatically in Fig. 3. Briefly, DNA restriction fragments containing 300–400 base pairs (bp) of 5' flanking sequence and either the first 12 codons (pEAB-1) or 67 codons (pEAB-2) of *RF2* were inserted into a restriction site polylinker in the 5' coding region of the *lacZ* gene encoded on plasmid p9AB1. The *lacZ* promoter was subsequently removed by *Bal31* digestion, and production of a fusion protein was confirmed by immunoprecipitation of 35 S-methionine-labelled protein using anti-RF2 antibodies (Fig. 4). Only the pEAB-2 fusion protein is immunoprecipitable with anti-RF2 antibodies; presumably the pEAB-1 product lacks a sufficient RF2 sequence to be recognized by the anti-RF2 antibodies.

Bacteria harbouring these gene fusions were assayed for β -galactosidase activity, and the results are shown in Table 1. Several conclusions can be drawn from these results: first, since a high-level activity is seen with the gene fusion lacking codon 26, we conclude that the stop codon may be the only means by which *RF2* expression is regulated, although we cannot exclude the possibility of an as yet undescribed 5' sequence, or sequences between codon 12 and 25, being involved in RF2 regulation.

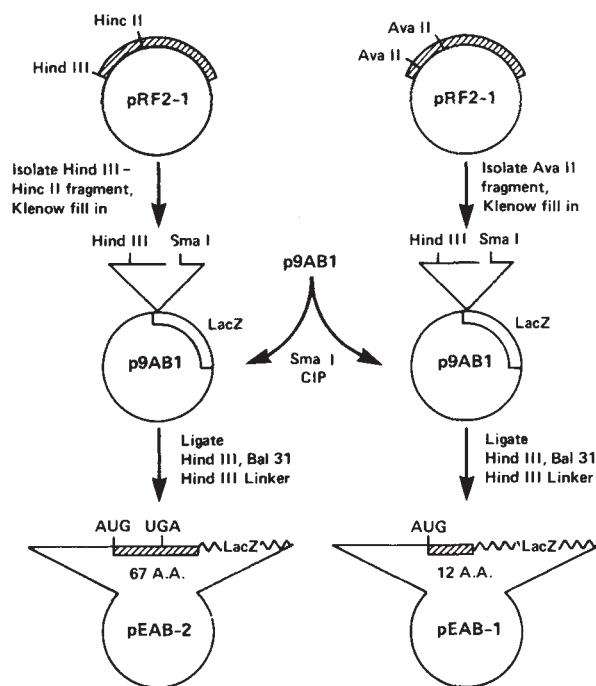


Fig. 3 The strategy for constructing *RF2-lacZ* fusion genes. The plasmid pRF2-1 contains the entire *RF2* gene with 400 bp of 5' flanking DNA. Two DNA fragments containing the *RF2* promoter, the initiating AUG, and either 12 codons or 67 codons of the coding region were isolated by restriction enzyme digestion and low melting agarose gel electrophoresis. Both fragments were made blunt-ended with the Klenow fragment of *E. coli* pol I, and ligated into *Sma*I-cut and dephosphorylated p9AB1. This plasmid contains the entire coding region of *lacZ* with a polylinker inserted in the amino-terminal end, and confers ampicillin resistance (provided by Bob Weiss, University of Utah). After ligation, the hybrid plasmid was redigested at a 5' flanking *Hind*III site, the adjacent *lac* promoter removed by *Bal31* digestion, and a *Hind*III linker inserted to assess the size of the deletion. Plasmids containing an equivalent amount of 5' flanking DNA from the *RF2* gene were selected for further study. The final plasmids are indicated at the bottom; pEAB-1 lacks the frameshift region, while pEAB-2 contains it.

This high level of activity also indicates that the *RF2* gene contains a strong promoter, as underscored by the threefold greater activity of pEAB-1 over the parent plasmid activity. Second, by comparing the β -galactosidase activity of each *RF2-lacZ* fusion, a rate of frameshifting across the *opal* stop codon can be determined. Since pEAB-2 exhibits 50% of the β -galactosidase activity of pEAB-1, in the presence of normal levels of endogenous RF2 the ribosome can be assumed to frameshift at this rate. Thus, without a compensatory suppression of translation by excess RF2, the region around the in-frame stop codon permits a remarkably high level of frameshifting.

Table 1 High-efficiency frameshifting

Plasmid	β -Galactosidase activity
pEAB-1	1.3×10^4
pEAB-2	6.4×10^3
p9AB1	4.5×10^3

β -Galactosidase activity of *RF2-lacZ* fusion proteins. The assay was performed, and the units of activity calculated, according to Miller¹⁷. Bacteria were grown in minimal A medium supplemented with amino acids. The host strain is Su1675, a *recA*⁻ derivative of CSH36 (Δ (*lac*pro), *Lac*⁻), provided by Bob Weiss. Plasmid copy number was determined by the method of Uhlin and Nordström¹⁸, and found to be essentially identical for all three plasmids.

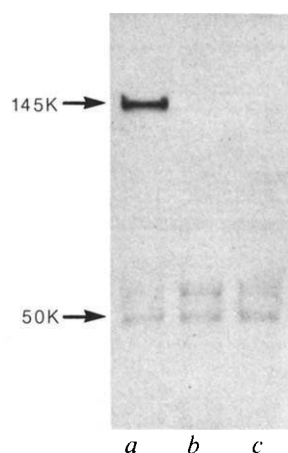


Fig. 4 Immunoprecipitation of RF2-LacZ fusion proteins. Exponentially growing *E. coli*, grown in minimal A medium¹⁷, was labelled with ³⁵S-methionine for 1 h and cell lysates made and immunoprecipitated with anti-RF2 antibodies, as described in Caskey *et al.*¹⁶, except that the proteins were electrophoresed through an 8% polyacrylamide gel. *a*, Immunoprecipitated proteins from cells containing pEAB-2; *b*, from cells containing pEAB-1; *c*, from cells containing the parent *lacZ* plasmid p9AB-1. The band seen in *a* at a relative molecular mass of 145K (145,000) represents the RF2-LacZ fusion protein. The plasmid pEAB-1 fails to make an immunoprecipitable fusion protein, probably due to insufficient RF2 sequences being present, while the LacZ protein encoded on p9AB-1 fails to react with anti-RF2 antibodies. The bands seen at the 50 K position correspond to the endogenous RF2 protein.

Several factors influence the rate of frameshifting. Codon context has been shown to be important in determining the capacity of a particular codon to frameshift, with certain codons and their cognate transfer RNAs being inherently more 'shifty', and particular surrounding sequences having marked effects on the rate of frameshifting⁶. In certain circumstances the identity of the base found 3' to a shifty codon seems to play an important part in determining the rate of frameshifting at that codon, perhaps by allowing codon/anticodon mismatch and ribosome slippage⁷. Codon context has also been shown to effect the efficiency of nonsense suppression⁸, missense suppression⁹ and *in vitro* misreading of UGU codons¹⁰. Since previous examples of frameshifting were uniformly low-frequency events, factors involved in the recognition of stop codons by RF2 may be still another factor contributing to this high-efficiency frameshifting.

Natural frameshifting is involved in the expression of several prokaryotic and eukaryotic genes. Translation of the major coat protein of phage T7 (ref. 11), the gag-pol fusion protein of Rous sarcoma virus¹², and the analogous overlapping open reading frames of the yeast Ty912 transposable element¹³ involve an obligate frameshift event, although the exact position of the frameshift in these genes is unknown. In some of the above examples frameshifting may provide a mechanism for maintaining a fixed ratio of two separate proteins. For RF2 it seems that the frameshift is solely a means of regulating RF2 translation.

Elaborating the mechanism of frameshifting requires, as a first step, identifying the neighbouring sequences required for high-level frameshifting. Since RF2 exhibits a very high rate of frameshifting relative to artificial frameshift mutations¹⁴, RF2 may provide a unique model system for refining the rules of frameshifting, and for studying the mechanism of stop codon recognition.

W.J.C. was supported by the Baylor College of Medicine Medical Scientist Training Program GM07330. This research was supported by NIH grant GM34438 and the Howard Hughes Medical Institute.

Received 24 March; accepted 13 May 1986.

1. Caskey, C. T. *Trends biochem. Sci.* **5**, 234-237 (1980).
2. Klein, H. A. & Capocchi, M. R. *J. biol. Chem.* **256**, 1055-1061 (1971).
3. Craigen, W. J., Cook, R. G., Tate, W. P. & Caskey, C. T. *Proc. natn. Acad. Sci. U.S.A.* **82**, 3616-3620 (1985).
4. Scolnick, E., Tompkins, R., Caskey, T. & Nirenberg, M. *Proc. natn. Acad. Sci. U.S.A.* **61**, 768-774 (1968).
5. Nomura, M., Jinks-Robertson, S. & Miura, A. in *Interaction of Translational and Transcriptional Controls in the Regulation of Gene Expression* (eds Grunberg-Manago, M. & Safer, B.) 91-104 (Elsevier, Amsterdam, 1982).
6. Weiss, R. *Proc. natn. Acad. Sci. U.S.A.* **81**, 5797-5801 (1984).
7. Weiss, R. & Gallant, J. *Genetics* (in the press).
8. Bossi, L. *J. molec. Biol.* **164**, 73-87 (1983).
9. Murgola, E. J., Pagel, F. T. & Hijazi, K. A. *J. molec. Biol.* **175**, 19-27 (1984).
10. Carrier, M. J. & Buckingham, R. H. *J. molec. Biol.* **175**, 29-38 (1984).
11. Dunn, J. J. & Studier, F. W. *J. molec. Biol.* **166**, 477-535 (1983).
12. Jacks, T. & Varmus, H. E. *Science* **230**, 1237-1242 (1985).
13. Clare, J. & Farabaugh, P. *Proc. natn. Acad. Sci. U.S.A.* **82**, 2829-2833 (1985).
14. Atkins, J. F., Elseviers, D. & Gorini, L. *Proc. natn. Acad. Sci. U.S.A.* **69**, 1192-1196 (1972).
15. Caskey, C. T., Scolnick, E., Tompkins, R., Milman, G. & Goldstein, J. *Meth. Enzym.* **20**, 367-375 (1971).
16. Caskey, C. T., Forrester, W. C., Tate, W. P. & Ward, C. J. *Bact.* **158**, 365-368 (1984).
17. Miller, J. H. *Experiments in Molecular Genetics*, 355 (Cold Spring Harbor Laboratory, New York, 1972).
18. Uhlin, B. E. & Nordström, K. *Plasmid* **1**, 1-8 (1977).

A mouse locus at which transcription from both DNA strands produces mRNAs complementary at their 3' ends

Trevor Williams & Mike Fried

Department of Tumour Virus Genetics, Imperial Cancer Research Fund, Lincoln's Inn Fields, London WC2A 3PX, UK

The organization and large size of the mammalian cell genome allows spatial separation of different transcription units. In those cases where more than one species of messenger are synthesized from the same cellular DNA sequence, they have been found to be generated from transcription proceeding in the same direction. These mRNAs always share regions of homology and can differ from one another as a result of differential processing (splicing and/or polyadenylation) or alternative initiation¹⁻¹⁴. In contrast, complementary mRNAs transcribed from opposite strands of the same cellular DNA sequence have not previously been observed. Here we have identified a region of mouse DNA at which processed mRNAs from two adjacent convergent transcription units overlap by 133 base pairs (bp) at their 3'-untranslated ends. One of the transcription units appears to encode a second mRNA which does not contain this overlapping region. This represents the first description of the natural occurrence of processed mammalian cell mRNAs transcribed from opposite strands of the same DNA sequence. The implications of these complementary regions in normal gene regulation are discussed in the context of the finding that the artificial introduction into cells of DNA constructs synthesizing anti-sense RNAs complementary to regions of mRNA transcribed from a chromosomal gene, can inhibit the gene's activity, presumably by the formation of double-stranded RNA¹⁵⁻¹⁹.

We have used the 'expression selection' technique to isolate cellular DNA sequences with enhancer-like properties (expression sequences) by virtue of their ability to reactivate a test gene lacking its own 5' regulatory sequences^{20,21}. A 19-kilobase (kb) *EcoRI* fragment of mouse cellular DNA containing the normal chromosomal location of the previously identified cellular expression sequence²⁰ has been cloned and characterized and found to encode a number of transcription units (T.W. and M.F., in preparation). The 746-bp *PvuII/EcoRI* subfragment (Fig. 1a) found at the right-hand end of the 19-kb clone is present as a single-copy sequence when used to probe Southern blots of mouse genomic DNA (Fig. 1b). In contrast, when this fragment is used to probe Northern blots of mouse poly(A)⁺ RNA, it hybridizes to two transcripts, of 1.2 and 3.0 kb (Fig. 1c). The presence of overlapping regions of the two

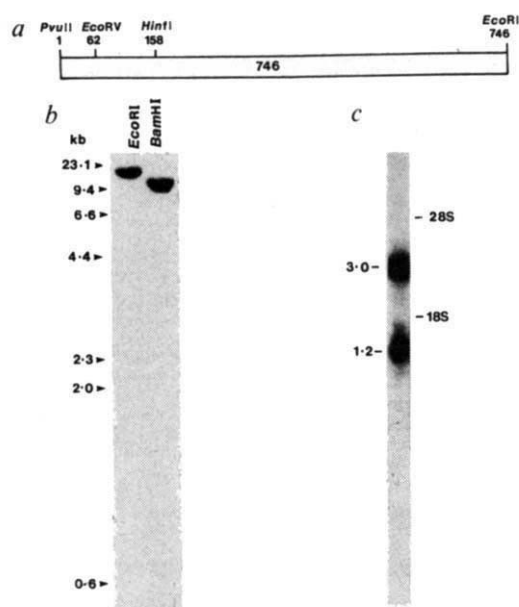


Fig. 1 Southern and Northern analysis of the 746-bp *PvuII*-*EcoRI* mouse cellular DNA fragment. *a*, Restriction map of the 746-bp *PvuII*-*EcoRI* fragment. This fragment was derived from one end of a 19-kb *EcoRI* clone (T.W. and M.F., in preparation) which contains a copy of a previously isolated mouse expression sequence²⁰, and which is subcloned into *EcoRI*/*SmaI*-cleaved pEMBL 8(+)³⁶. The *PvuII*, *EcoRV*, *HinfI* and *EcoRI* sites are indicated. *b*, Detection of a single copy of the 746-bp sequence in the mouse genome. A Southern blot of mouse genomic DNA digested with either *EcoRI* or *BamHI* was hybridized to the 746-bp probe. The numbers on the left show the sizes of the DNA markers (in kb). *c*, Northern blotting analysis of BALB/c TS-A-3T3 poly(A)⁺ RNA hybridized to the 746-bp fragment identifies two distinct mRNA species of 3.0 and 1.2 kb. On the right are shown the positions of the 28S and 18S ribosomal RNA markers.

Methods. For Southern blotting, 16 μ g of BALB/c 3T3 DNA was digested with either *EcoRI* or *BamHI*, fractionated by electrophoresis on a 0.8% agarose gel, and transferred to nitrocellulose as described previously³⁷. For Northern blotting, total cellular RNA was prepared from BALB/c TS-A-3T3 cells²⁵ by the guanidinium/CsCl method³⁸ and poly(A) selected on oligo(dT)-cellulose; 5 μ g of poly(A)⁺ RNA was fractionated on a 1% agarose/formaldehyde gel and transferred to nitrocellulose after partial alkaline hydrolysis³⁸. Hybridization conditions were identical for Northern and Southern blotting and both types of blot were washed to a stringency of $0.5\times$ SSC at 68°C essentially as described previously³⁷.

mRNAs was initially suggested by the finding that when uniformly labelled single-stranded probes specific for either strand of the 746-bp fragment were annealed to poly(A)⁺ RNA, both strands produced *S*₁-nuclease-resistant products, and the sum of these products was greater than 746 nucleotides. To more accurately localize the transcripts and determine the extent of their overlap, 3' *S*₁-nuclease mapping was performed using either a 746-nucleotide probe 3'-labelled at the *EcoRI* site or a 684-nucleotide probe 3'-labelled at the *EcoRV* site at position 62 (Fig. 2*a*, probes A and B respectively). The probes were annealed to mouse RNA, and the *S*₁-nuclease-resistant products fractionated on polyacrylamide gels. Probe A, homologous to the top (A) strand, protects a 500-bp fragment (Fig. 2*b*, lane 3), locating a splice donor or polyadenylation site around nucleotide 242. Probe B, homologous to the bottom (B) strand, mainly protects a 310-bp fragment (Fig. 2*c*, lane 3), locating a splice donor or polyadenylation site around residue 372, but on the opposite strand. This indicates that the transcripts are derived from opposite strands and overlap by ~130 nucleotides.

To analyse the overlapping region in detail, complementary DNA clones corresponding to both transcripts were isolated

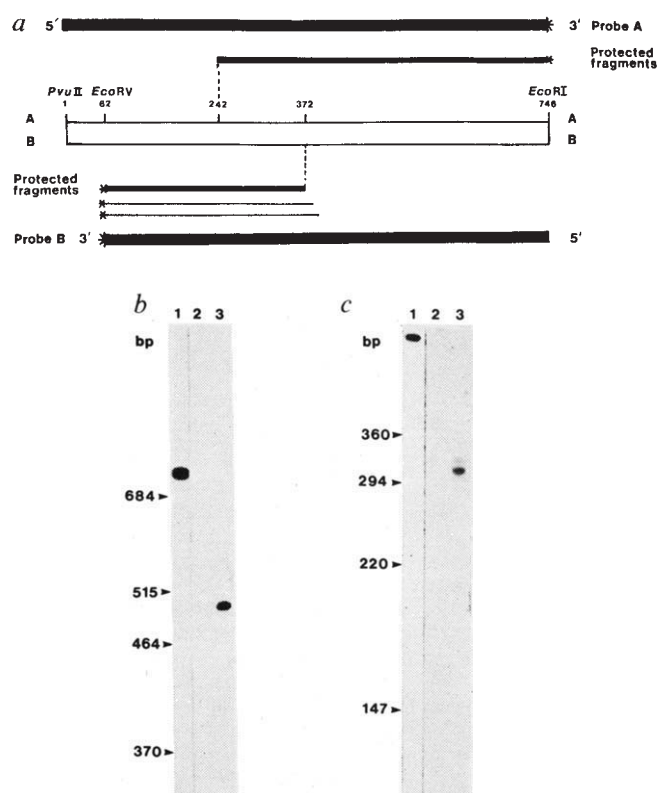


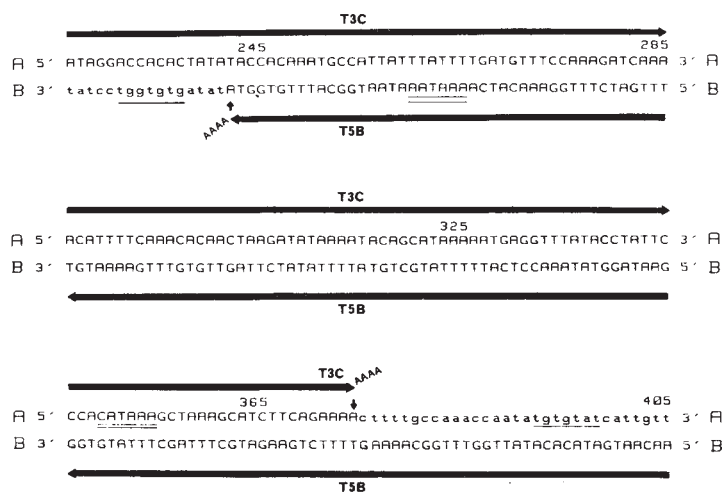
Fig. 2 Identification by 3' *S*₁ mapping of a region of mouse DNA where convergent transcription units overlap. *a*, Diagrammatic representation of the probes used for 3' *S*₁ mapping and the overlapping protected fragments. The open box represents the 746-bp fragment, with the A strand at the top and the B strand at the bottom. Nucleotide (nt) numbers above the DNA fragment are given from the *PvuII* site. The 746-nt probe A is illustrated by the solid box at the top, and is 3'-end-labelled at the *EcoRI* site (indicated by *). The 684-nt probe B, 3'-end-labelled at the *EcoRV* site, is similarly shown at the bottom. The major fragments protected in *b* and *c* are represented by the thick lines between the probes and the genomic sequence; the minor bands in *c* are represented similarly by thin lines. *b*, 3' *S*₁ mapping using the 746-nt probe A homologous to the A strand (see *a*). Lane 1, untreated probe; lane 2, probe plus 40 μ g yeast RNA; lane 3, probe plus 40 μ g BALB/c TS-A-3T3 total RNA. Marker DNA sizes are given on the left (in bp). *c*, 3' *S*₁ mapping using the 684-nt probe B homologous to the B strand (see *a*). Lane 1, untreated probe; lane 2, probe plus 40 μ g yeast RNA; lane 3, probe plus 40 μ g BALB/c TS-A-3T3 total RNA. Marker DNA sizes are given on the left (in bp).

Methods. The 746-bp fragment subcloned in pEMBL 8(+) was cleaved with either *EcoRI* (for probe A) or *EcoRV* (for probe B). Probe A was 3'-end-labelled at the *EcoRI* site using Klenow polymerase and ³²P-dATP and re-cut at the *BamHI* site present in the pEMBL 8(+) polylinker sequence. Probe B was 3'-end-labelled at the *EcoRV* site using T₄ DNA polymerase and ³²P-TTP and re-cut with *EcoRI*. Probes A and B were then prepared by electroelution after fractionation on 1% agarose gels. Procedures for RNA/DNA hybridization and *S*₁-nuclease digestion were essentially as described elsewhere³⁹. *S*₁-protected products were analysed by electrophoresis on 4% (*b*) or 6% (*c*) polyacrylamide gels containing 50% urea.

from a mouse cell cDNA library and sequenced. Figure 3 shows the relationship of the cDNAs T3C and T5B to the sequence of the 746-bp genomic fragment. The 3' 374 nucleotides of T3C (up to the start of the poly(A) tract) are homologous to the genomic sequence of the A strand from nucleotide 1 to 374. In addition, the 3' 505 nucleotides of T5B (up to the start of the poly(A) tract) are homologous to the genomic sequence of the B strand from position 746 to 242 (nucleotide numbers corre-

Fig. 3 A comparison of the genomic sequence of the 746-bp fragment with the cDNA sequences of T3C and T5B in the region of the overlap. The extent and orientation of the cDNAs are shown by the thick, solid arrowed lines above or below the sequence. Upper-case letters indicate sequences present in the processed poly(A)⁺ RNA; lower-case letters denote sequences 3' of the poly(A) addition sites. The vertical arrows indicate the last nucleotide of the genomic sequence which is homologous to the sequence of either T3C (A strand) or T5B (B strand). Beyond these positions the presence of poly(A) tracts in the cDNAs distinguishes them from the genomic sequence. The postulated poly(A) addition signals for both transcripts are indicated by double underlining. The postulated poly(A) addition signal for transcripts homologous to the B strand is the canonical AAUAAA⁴⁰, while that for transcripts homologous to the A strand, which may display microheterogeneity in the nucleotide used for poly(A) addition, is the much rarer CAUAAA⁴¹. The TG nucleotide blocks seen downstream of many poly(A) addition sites are underlined⁴²⁻⁴⁵. In contrast, the consensus sequence CAYUG, postulated to be important in the 3' processing of primary transcripts⁴⁶, is not seen in the vicinity of the poly(A) addition site for either transcript.

Methods. A cDNA library was constructed from BALB/c TS-A-3T3 poly(A)⁺ RNA as described elsewhere⁴⁷, with the following modifications: after second-strand synthesis the cDNA was repaired with T₄ DNA polymerase and ligated into λ gt10 (ref. 48) after the addition of *Eco*RI linkers. cDNAs T3C and T5B were isolated on the basis of their homology to the 746-bp genomic fragment, and their *Eco*RI cDNA inserts were subcloned into pAT153. The sequence of the 746-bp fragment and the cDNA clones T3C and T5B in the region of the overlap was obtained for both strands by the dideoxy method⁴⁹ after appropriate restriction fragments had been subcloned into either M13mp18 or M13mp19 (ref. 50).



spond to the A strand sequence in Fig. 2). Thus, the two cDNAs analysed here are derived from mRNAs transcribed from opposite strands of the 746-bp genomic fragment (T3C from the B-strand template and T5B from the A-strand template) and overlap by 133 nucleotides at their 3' ends (Fig. 3), which is in agreement with the 3' S₁ mapping data presented in Fig. 2. The minor bands observed using probe B (Fig. 2c, lane 3) may be due to microheterogeneity in the choice of site for poly(A) addition²²⁻²⁴, in which case the exact extent of the overlap for processed mRNA would depend on which nucleotide was used for poly(A) addition for mRNAs homologous to the A strand (corresponding to cDNA T3C).

To determine which of the poly(A)⁺ RNA species in Fig. 1c corresponded to which cDNA clone, a truncated fragment of T3C (deleted 3' to the *Hinf*I site; see Fig. 4e) lacking the 3' overlapping sequences was used as a probe for Northern blotting and found to hybridize only to the 1.2-kb mRNA species (Fig. 4, compare a and b). When the cDNA clone RMC4, which is homologous to T5B but is truncated at its 3' end and contains only 18 bp of the overlapping sequences (Fig. 4c,e), was used as a probe for Northern blotting, it hybridized to 3.0-kb mRNA, but not to the 1.2-kb mRNA (Fig. 4c). Probe RMC4 also hybridized to a 2.4-kb mRNA (Fig. 4c,e), which was not visible if the 746-bp genomic clone was used as a probe (Fig. 4b). However, when used to probe Southern blots of mouse genomic DNA, clone RMC4 was found to represent single-copy sequence in the genome (data not shown); this suggests that, in addition to the transcripts homologous to the B strand which have been polyadenylated in the 746-bp fragment, typified by T5B (see Figs 3, 4e), there is an alternative shorter transcript terminating before the overlapping region of the 746-bp genomic sequence is reached (Fig. 4e). Further experiments will be necessary to determine whether the 2.4-kb poly(A)⁺ species is derived from this transcription unit by alternative 3' exon usage or alternative polyadenylation. We are presently cloning the mouse genomic DNA adjacent to the 746-bp fragment in order to further analyse the relationship of the 2.4-kb mRNA to the 3.0-kb mRNA.

The presence of the 1.2-kb and 3.0-kb mRNAs was assessed by Northern blotting in several normal and transformed clonal cell lines (Fig. 4d). No differences were detected in the ratios of the steady-state levels of the two mRNAs between normal BALB/c 3T3 cells and their transformed derivative BALB/c

TS-A-3T3 cells²⁵. However, the ratios of the steady-state levels of the two mRNAs in F9 (ref. 26) and MH15 (ref. 27) teratocarcinoma cells differed from those found in differentiated mouse cells (Fig. 4d). In the undifferentiated cells, the 3.0-kb mRNA appears to be down-regulated with respect to the 1.2-kb mRNA.

The function of the 1.2-, 2.4- and 3.0-kb mRNAs remains to be determined. However, in this context, the 5' end of the gene encoding the 1.2-kb mRNA maps in close proximity to the expression sequence²⁰ and displays a number of properties which are normally associated with housekeeping genes (T.W. and M.F., in preparation). The structures of the 3.0-kb and 2.4-kb mRNAs are presently being characterized.

The overlapping transcripts described here raise several interesting questions concerning the organization of genetic information and the control of gene expression. The mammalian genome characteristically contains individual genes separated by large regions of untranscribed DNA. In the present report we have identified a mouse genetic locus at which two processed poly(A)⁺ RNA species transcribed from opposite strands overlap by 133 nucleotides at their 3' ends. This finding implies a complex sequence organization in this region as the poly(A) addition signal for one transcript is contained in the exon sequence of the other, and vice versa. In addition, for those vertebrate genes in which it has been studied, transcription does not terminate at the poly(A) addition site but continues for some distance downstream. Therefore, the sequences that may be involved in the transcriptional termination or processing of one transcript are also probably encoded in the sequence giving rise to the transcript from the other strand, and vice versa. Therefore, it may be significant that the overlap is in the 3'-untranslated region of both mRNAs. Such overlapping transcripts have been identified previously in mammalian viruses, for example, the early and late mRNAs of polyoma virus overlap by 50 nucleotides and those of simian virus 40 by 87/88 nucleotides, although there is temporal control of the expression of these opposite strands²⁸.

In prokaryotes the formation of double-stranded RNA by the hybridization of complementary transcripts is important in the regulation of gene expression and replication. Such a mechanism has been shown to be responsible for the inhibition of translation of the *ompF* (ref. 29) and IS10 transposase³⁰ genes, and the control of replication of the plasmid ColE1 by inhibiting RNA

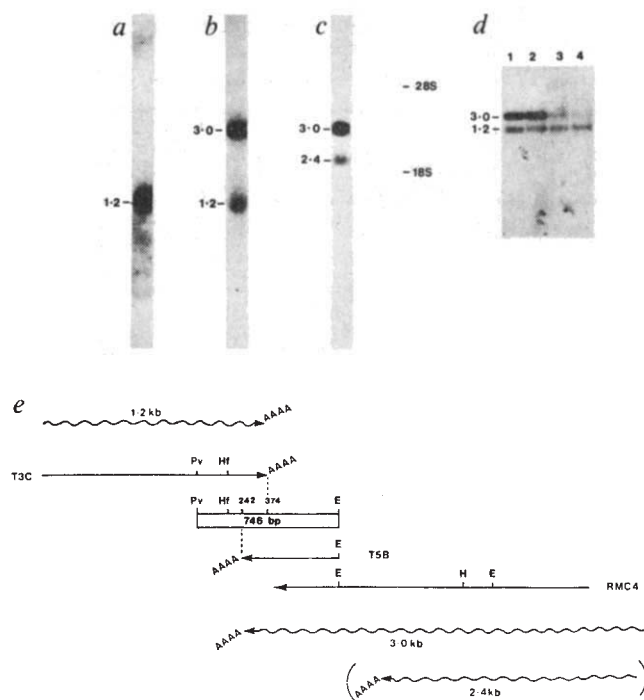


Fig. 4 Analysis of mRNAs homologous to the overlapping region.

a, Northern blot analysis of BALB/c TS-A-3T3 poly(A)⁺ RNA to determine which mRNA is homologous to cDNA clone T3C (see **e**). A fragment of T3C lacking sequences 3' of the *HinfI* site (nucleotide 158 on the genomic sequence shown in **e**) and so missing the overlapping sequence, was used as a probe. **b**, Northern blot analysis of BALB/c TS-A-3T3 poly(A)⁺ RNA hybridized to the 746-bp fragment. **c**, Northern analysis of BALB/c TS-A-3T3 poly(A)⁺ RNA hybridized to cDNA RMC4 (see **e**) identifies mRNAs of 3.0 kb and 2.4 kb. cDNA RMC4 was isolated from a cDNA library derived from mouse mammary tumour virus-transformed mouse cells (given by R. Moore) on the basis of its homology to the 746-bp genomic fragment. RMC4 is truncated at its 3' end 115 bp before the poly(A) addition site and so lacks all but 18 bp of the overlapping sequences that would hybridize to the 1.2-kb mRNA. **d**, Northern blot analysis of mouse poly(A)⁺ RNA from differentiated cell lines (lane 1, BALB/c TS-A-3T3; lane 2, BALB/c 3T3) and undifferentiated teratocarcinoma cell lines (lane 3, MH15; lane 4, F9) hybridized to the 746-bp fragment. **e**, Organization of cDNA clones and transcripts in the vicinity of the 746-bp fragment. The 746-bp cellular DNA sequence is indicated by an open box. The straight lines represent the cDNA clones, with the arrows showing the direction of transcription and AAAA indicating the presence of a poly(A) tail. The dotted lines between the 746-bp fragment and T3C or T5B indicate the extent of the cDNAs with respect to the sequence of the genomic A strand; the last nucleotide of each cDNA is given above the 746-bp fragment (see Fig. 3). Wavy lines indicate the positions of the mRNAs, with their direction of transcription shown by arrows, and the poly(A) tails indicated by AAAA. The wavy line in parentheses represents the 2.4-kb mRNA which is related to the cDNA clone RMC4 but which terminates before the *EcoRI* site is reached (see **c**). The positions of relevant restriction enzyme sites are indicated: E, *EcoRI*; Pv, *PvuII*; H, *HindIII*; and Hf, *HinfI*. RNA preparation and Northern blotting procedures were identical to those described in Fig. 1 legend.

priming³¹⁻³³. In eukaryotes the introduction of anti-sense RNA, or vectors expressing anti-sense RNA, into cells has been used as a tool for the inhibition of gene product expression^{15,16}. In a cytoplasmic location, anti-sense RNAs which are complementary to the 5' sequence of a mRNA have been found to be more inhibitory than anti-sense RNAs complementary to the 3' sequence. This effect appears to be mediated by the ability of anti-sense RNAs to inhibit the translation of the specific mRNA

to which they are complementary^{17,18}. However, when anti-sense transcripts are expressed in a nuclear location, inhibition probably occurs via the formation of double-stranded RNA, which prevents the transport of sense transcripts from the nucleus; in this case, sequences complementary to the 3' portion of the sense transcript are as potent as those complementary to the 5' sequences¹⁹. In addition, in *Saccharomyces cerevisiae* the *CYC1* gene and an adjacent gene are convergently transcribed from opposite strands, but transcription termination signals located between the two genes normally ensure that transcripts from one gene do not read-through into the other gene and vice versa. The deletion of these termination signals generates overlapping convergent transcription from opposite strands and has a marked deleterious effect on the mRNA levels yielded from both transcription units³⁴.

Therefore, the two overlapping transcripts described here may display some novel mechanism of control of mammalian gene expression. First, if the overlapping transcripts were expressed at the same time, it might be expected that RNA polymerase complexes travelling in opposite directions along a DNA duplex might encounter a considerable degree of steric hindrance, especially as transcription probably continues past the poly(A) addition site³⁵. Second, although the abundance of the mRNAs is fairly low (<0.01% for each of the two mRNA species), the local concentration of the two opposite-strand transcripts in the vicinity of the 3'-ends of the transcription units might favour RNA duplex formation with concomitant inhibition of RNA processing and/or transport from the nucleus. Thus the two strands might be only rarely transcribed and their mRNAs very stable, or there might be some temporal or spatial control of expression such that their transcription occurs at different times in the cell cycle or from different alleles. Alternatively, coincident convergent transcription might be involved in controlling the expression of one or both of these transcripts. For example, it may be that when transcription of the 1.2-kb mRNA species is required, the shorter, non-overlapping 2.4-kb mRNA is generated from the opposite strand. Finally, an effect on the translation of either the 1.2-kb or the 3.0-kb mRNA by the overlap cannot be excluded. Further studies on the transcription of this mouse genomic locus may answer some of these questions.

We thank Drs P. Goodfellow, N. Jones, I. Kerr, P. Nurse and M. Parker for helpful discussions, and Mrs G. Briody and Mrs C. Middlemiss for typing the manuscript.

Received 28 February; accepted 15 April 1986.

- Alt, F. W. *et al. Cell* **20**, 293-301 (1980).
- Early, P. *et al. Cell* **20**, 313-319 (1980).
- Rosenfeld, M. G. *et al. Nature* **304**, 129-135 (1983).
- Medford, R. M., Nguyen, H. T., Destree, A. T., Summers, E. & Nadal-Ginard, B. *Cell* **38**, 409-421 (1984).
- Kornblihtt, A. R., Umezawa, K., Vibe-Pedersen, K. & Baralle, E. *EMBO J.* **4**, 1755-1759 (1985).
- de Ferri, F. *et al. Cell* **43**, 721-727 (1985).
- Leonard, W. J. *et al. Science* **230**, 633-639 (1985).
- Setzer, D. R., McGrogan, M., Hunberg, J. H. & Schimke, R. T. *Cell* **22**, 361-370 (1980).
- Tosi, M., Young, R. A., Hagenbuchle, O. & Schibler, U. *Nucleic Acids Res.* **9**, 2313-2323 (1981).
- Setzer, D. R., McGrogan, M. & Schimke, R. T. *J. Biol. Chem.* **257**, 5143-5147 (1982).
- Parnes, J. R. & Robinson, R. R. *Nature* **302**, 449-452 (1983).
- Sakakibara, M., Mukai, T., Yatsuki, H. & Hori, K. *Nucleic Acids Res.* **13**, 5055-5069 (1985).
- Müller, R. J. *et al. Cell* **42**, 931-939 (1985).
- Schibler, U., Hagenbuchle, O., Wellauer, P. K. & Pittet, A. C. *Cell* **33**, 501-508 (1983).
- Izant, J. G. & Weintraub, H. *Cell* **36**, 1007-1015 (1984).
- Izant, J. G. & Weintraub, H. *Science* **229**, 345-352 (1985).
- Melton, D. A. *Proc. natn. Acad. Sci. U.S.A.* **82**, 144-148 (1985).
- Harland, R. & Weintraub, H. *J. Cell Biol.* **101**, 1094-1099 (1985).
- Kim, S. K. & Wold, B. *Cell* **42**, 129-138 (1985).
- Fried, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 2117-2121 (1983).
- Ford, M., Davies, B., Griffiths, M., Wilson, J. & Fried, M. *Proc. natn. Acad. Sci. U.S.A.* **82**, 3370-3374 (1985).
- Sasavage, N. L., Smith, M., Gillam, S., Woychik, R. O. & Rottman, F. M. *Proc. natn. Acad. Sci. U.S.A.* **79**, 223-227 (1982).
- Wiedemann, L. M. & Perry, R. P. *Molec. cell. Biol.* **4**, 2518-2528 (1984).
- Simonsen, C. C. & Levinson, A. D. *Molec. cell. Biol.* **3**, 2250-2258 (1983).
- Ruley, H. E., Lania, L., Chaudry, F. & Fried, M. *Nucleic Acids Res.* **10**, 4515-4524 (1982).
- Bernstine, E. G., Hooper, M. L., Grandchamp, S. & Ephrussi, B. *Proc. natn. Acad. Sci. U.S.A.* **70**, 3899-3903 (1973).
- Solter, D. & Knowles, B. *Proc. natn. Acad. Sci. U.S.A.* **75**, 5565-5569 (1978).

28. Tooze, J. (ed.) *DNA Tumour Viruses* 2nd edn (Cold Spring Harbor Laboratory, New York, 1981).
29. Mizuno, T., Chou, M.-Y. & Inouye, M. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1966-1970 (1984).
30. Simons, R. W. & Kleckner, N. *Cell* **34**, 683-691 (1983).
31. Tomizawa, T.-I., Itoh, T., Selzer, G. & Som, T. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1421-1425 (1981).
32. Cesareni, M., Muesing, M. A. & Polisky, B. *Proc. natn. Acad. Sci. U.S.A.* **79**, 6313-6317 (1982).
33. Tomizawa, J.-I. & Itoh, T. *Cell* **31**, 575-583 (1982).
34. Zaret, K. S. & Sherman, F. *Cell* **28**, 563-573 (1982).
35. Birnstiel, M. L., Busslinger, M. & Strub, K. *Cell* **41**, 349-359 (1985).
36. Dente, L., Cesareni, G. & Cortese, R. *Nucleic Acids Res.* **11**, 1645-1655 (1983).
37. Hayday, A., Ruley, H. E. & Fried, M. J. *J. Virol.* **44**, 67-77 (1982).
38. Maniatis, T., Fritsch, E. F. & Sambrook, J. in *Molecular Cloning: a Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
39. Favaloro, J., Triesman, R. & Kamen, R. *Meth. Enzym.* **65**, 718-749 (1980).
40. Proudfoot, N. J. & Brownlee, G. G. *Nature* **263**, 211-214 (1976).
41. Wickens, M. & Stephenson, P. *Science* **226**, 1045-1051 (1984).
42. Taya, T. et al. *EMBO J.* **1**, 953-958 (1982).
43. Woychik, R. P., Lyons, R. H., Post, L. & Rottman, F. M. *Proc. natn. Acad. Sci. U.S.A.* **81**, 3944-3948 (1984).
44. Gil, A. & Proudfoot, N. J. *Nature* **312**, 473-474 (1984).
45. McLauchlan, J., Gaffney, D., Whitton, J. L. & Clements, J. B. *Nucleic Acids Res.* **13**, 1347-1368 (1985).
46. Berget, S. M. *Nature* **309**, 179-182 (1984).
47. Gubler, U. & Hoffman, B. J. *Gene* **25**, 263-269 (1983).
48. Huynh, T. V., Young, R. A. & Davis, R. W. in *DNA Cloning: a Practical Approach* Vol. 1 (ed. Glover, D. M.) 49-78 (IRL, Oxford, 1985).
49. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
50. Yanisch-Perron, C., Vieira, J. & Messing, J. *Gene* **33**, 103-119 (1985).

Overlapping transcription units in the dopa decarboxylase region of *Drosophila*

Charlotte A. Spencer, R. Daniel Gietz
& Ross B. Hodgetts

Department of Genetics, University of Alberta, Edmonton, Alberta,
Canada T6G 2E9

The many examples of overlap in the genes of various viruses and bacteria illustrate that the parsimonious utilization of the coding capacity of DNA is relatively common amongst prokaryotes (for review, see ref. 1). The recent discovery of a pupal cuticle gene within an intron of the completely unrelated *Gart* locus in *Drosophila*² shows that overlapping transcription units also exist in higher organisms. However, the prevalence of such phenomena is unknown. We report here a quite different situation of overlap between the 3' termini of a pair of convergent transcription units in another region of the *Drosophila* genome. This 88-base-pair (bp) genomic region encodes the 3' terminus of the messenger RNA for the enzyme dopa decarboxylase (*Ddc*) and, in opposite orientation, the 3' terminus of the adjacent gene whose function is unknown. An analysis of the temporal and spatial distribution of the two transcripts within the organism shows that high levels of both transcripts are never concordant. However, within the testes, where the 3' transcript is maximally expressed, low levels of *Ddc* transcript were detected. This result raises the possibility that a hybrid molecule involving the two transcripts forms *in vivo* or that transcription interference occurs, with concomitant regulatory implications.

Previous work has shown that the dopa decarboxylase region of *Drosophila*, as defined by the deficiency *DF(2L)TW130* (ref. 3), contains a cluster of 18 genes, 14 of which appear to function in cuticle development and catecholamine metabolism⁴. Four genes, including *Ddc*, have been mapped within a 12-kilobase (kb) region and primary transcripts from these genes account for ~10.5 kb of the genomic DNA⁵. This tight clustering of genes has complicated our previous analysis of *Ddc* transcripts⁶, because the DNA probes we used in that study were found to encompass both *Ddc* and its 3'-adjacent transcription unit⁷. Strand-specific RNA probes revealed that the 3'-adjacent transcript originated on the opposite strand from that of the *Ddc* transcript and a hybridization of Northern blots to a series of nested single-stranded probes showed that the 3' termini of the

Ddc gene and the adjacent gene apparently lie on opposite sides of the *HpaI* site shown in Fig. 1. However, subsequent analysis of the DNA sequence in this region showed that putative polyadenylation signals for the two genes are present on opposite sides of the *HpaI* site (see Fig. 1). This result suggested the possibility of a slight amount of overlap between the two transcripts in the vicinity of the *HpaI* site. We have now obtained complementary DNA clones for both *Ddc* and the 3'-flanking gene, and a comparison of their DNA sequences with the genomic DNA sequence confirms that an 88-bp overlap between the two 3' termini exists.

A cDNA library in λ gt10 was made using poly(A)⁺ RNA from newly eclosed adults as described elsewhere⁸. Two clones each of *Ddc* and the 3'-flanking gene were chosen for further analysis. The inserts, which ranged in size from 1.1 to 2.0 kb, were re-cloned into M13mp19 and the ends of each fragment were sequenced (Fig. 2). Clones ACP, ACQ and ACR represented intact 3' termini, as revealed by tracts of oligo(dT) (or oligo(dA), depending on their orientation within the cloning vector) not found in the genomic sequence. The polyadenylation site of the 3'-adjacent gene was assigned unambiguously from the two cDNA clones of different length (ACQ and ACR); it is indicated by an asterisk beneath the genomic sequence in Fig. 2. A canonical polyadenylation signal sequence, 5'AATAAA3', lies 16 bases upstream of the terminal C residue (boxed in Fig. 2). The polyadenylation site for the *Ddc* gene (indicated by an asterisk above the genomic sequence in Fig. 2) could not be assigned unambiguously from the cDNA sequence because of

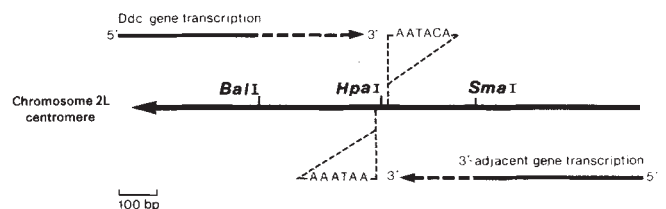


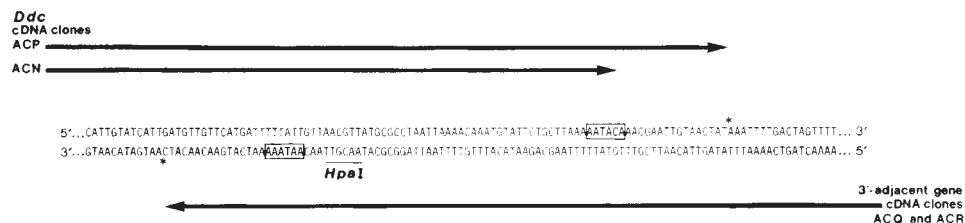
Fig. 1 Mapping the 3' termini of *Ddc* and the adjacent gene. The figure summarizes the results of our earlier work in which strand-specific RNA probes were used in a Northern analysis of poly(A)⁺ RNA from the mature larval and young adult stages of *Drosophila*⁷. Single-stranded *Ddc*-specific probes indicated that transcription of this gene terminated in the *BclI*-*HpaI* interval, while probes specific for the flanking gene indicated that its transcripts terminated in the *HpaI*-*SmaI* interval. Possible polyadenylation signals, chosen on the basis of their similarity to the eukaryotic consensus sequence 5'AATAAA3' (see ref. 9), were obtained from genomic DNA sequence analysis. Their precise locations with respect to the *HpaI* site are shown in Fig. 2.

the uncertainty caused by a run of adenine residues in the genomic sequence. However, we believe the site chosen at the T residue in Fig. 2 is the most likely one, based on its similarity to the eukaryotic polyadenylation addition sequence 5'...PyA...3' (ref. 9); a hexameric sequence (5'ATTACA3') which could control poly(A)⁺ addition at this residue lies 16 bases upstream of it. Although this sequence appears to be an inefficient signal for transcription termination¹⁰, the other possible hexamers 5'AACAA3' and 5'AATTAA3' lie 38 and 43 bases upstream, farther than any reported in the most recent review⁹.

The extent of overlap that we have found between the 3' termini of the poly(A)⁺ mRNAs may reflect an even greater overlap between the unprocessed, non-adenylated transcripts, assuming that transcription termination occurs in *Drosophila* 1 kb or farther downstream of the poly(A)⁺ signal sequence, as is known to occur in other systems⁹. The overlap also raises the intriguing possibility that a sense/anti-sense hybrid might exist *in vivo*, with regulatory implications for the expression of one

Fig. 2 Genomic DNA sequence in the region of overlap between the *Ddc* transcription unit and the 3'-flanking gene. The sequence of both strands of the genomic DNA was determined by the dideoxynucleotide method of Sanger¹⁵ and constitutes part of a 5-kb sequence including the entire *Ddc* gene⁸. Poly(A) signal sequences are boxed. **Methods.** Sequences of the cDNA

molecules near their 3' termini were obtained from single- or double-stranded¹⁶ templates in which the cDNA constructs were cloned into M13mp19. The cDNA libraries were prepared from poly(A)⁺ adult RNA as described elsewhere⁸. *Ddc*-specific clones were identified using a nick-translated *Hpa*I genomic fragment which extends 3.6 kb upstream of the *Hpa*I site shown in Fig. 1. Although this fragment actually includes the 3'-terminal 26 nucleotides of the flanking gene, the extent of homology between the probe and the 3' gene is insufficient to generate a stable hybrid under the conditions used. Clones of the 3'-flanking transcription unit were identified using the strand-specific probe 10 (ref. 7), which is homologous to the *Hpa*I-*Sma*I interval shown in Fig. 1.



or both mRNAs. We therefore undertook analyses of the temporal and spatial distribution of *Ddc* transcripts and of the 3'-adjacent transcripts during the life cycle of *Drosophila*. Our earlier work indicates that the developmental expression of the two genes is quite different. Maximum levels of dopa decarboxylase occur in 18-h embryos, at each larval moult, at pupariation and at adult eclosion¹¹. The 3'-adjacent transcript, however, is abundant only in adult stages and in 1-4-h embryos⁷. These data make it unlikely that the region of overlap between the two transcripts has biological significance during embryonic or larval life. However, both transcripts are present in adult flies (Fig. 3). Between 0 and 4 days, the level of mature *Ddc* transcript (the 2.0-kb species^{6,12}) decreases, although substantial amounts of the 3.0-kb *Ddc* species are still present (Fig. 3a). High levels of the 3'-adjacent transcript are present in both newly eclosed and 4-day-old adults and an obvious sexual dimorphism in both the amount and size distribution of this transcript is apparent (Fig. 3b).

To further define the distribution of the two transcripts within adults, we prepared RNAs from several dissected fractions of newly eclosed adults and quantitated the two transcripts by dot hybridization (Fig. 4). Control experiments (not shown) indi-

cated no detectable hybridization to either of the strand-specific probes in purified ribosomal or poly(A)⁻ cytoplasmic RNA from young adults. The pattern of *Ddc* expression in whole heads, thoraces and abdomens of both males and females (Fig. 4a) is consistent with a distribution of the enzyme in epidermal¹³ and neural tissue¹⁴. The higher concentration of *Ddc* transcripts in male abdomens (Fig. 4a, dot 7) than in female abdomens (a, dot 3) probably results from the contribution of testicular *Ddc* transcripts (a, dot 8) to the male abdominal sample. Longer exposures indicated a very low but detectable level of *Ddc* transcript in dissected ovaries (Fig. 4a, dot 4).

The pattern observed for the 3'-adjacent transcripts appears almost complementary to that of the *Ddc* transcripts (Fig. 4b): low, but detectable levels of the former transcript occur in both male and female heads and thoraces, and in female abdomens and ovaries. Clearly, though, the 3'-adjacent gene transcripts accumulate primarily in the testes (Fig. 4b, dot 8). A Northern analysis of total RNA samples from the dissected tissue samples confirmed that the male-specific size distribution of the 3'-adjacent gene transcripts (Fig. 3b) is also seen in the testes. Similarly, the female-specific size distribution (Fig. 3b) is identical to that contributed by ovaries and 2-h embryos (data not shown).

The data presented here, and our previous work⁷, indicate that high levels of the two overlapping transcripts are found only within young adults. However, the maximal steady-state levels of the two transcripts occur in different tissues. *Ddc* transcripts are confined primarily to the epidermis of males and females whereas transcripts of the 3'-adjacent gene accumulate primarily in the testes. The presence of low levels of *Ddc* transcript in the testes makes this the most likely place for formation

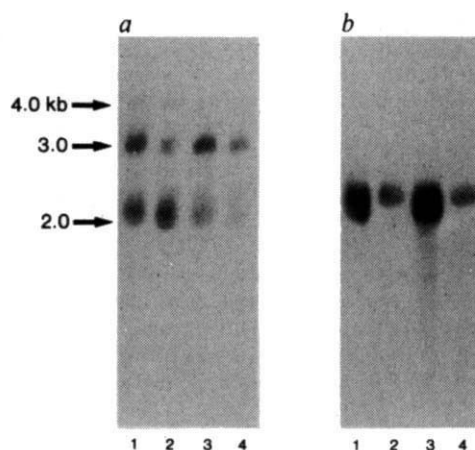


Fig. 3 Northern analysis of transcript levels in male and female adults. Duplicate samples of poly(A)⁺ RNA (5 µg) were subjected to Northern analysis as described previously⁷. The autoradiogram in *a* shows the hybridization response to the *Ddc*-specific RNA probe 5 prepared by *in vitro* transcription of a 1.5-kb genomic fragment including ~600 bp from the 3' end of *Ddc*⁷. *b*, The response to probe 4 (ref. 7), transcribed from the opposite strand of the same genomic fragment and therefore complementary to transcripts of the 3'-adjacent gene. Lanes 1, newly eclosed male; 2, newly eclosed female; 3, 4-day-old adult male; 4, 4-day-old adult female.

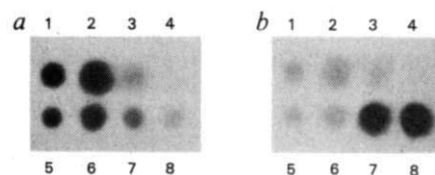


Fig. 4 Tissue distribution of transcripts of *Ddc* and the 3'-adjacent gene. RNA dot-blots were hybridized with: *a*, RNA probe 12, which is homologous to only RNA molecules transcribed from the 3'-flanking gene⁷; and *b*, RNA probe 5, which is homologous to only *Ddc* transcripts⁷. Total RNA was extracted as described previously from newly eclosed animals that had been dissected in D-20 medium¹⁷. Twelve animal equivalents were spotted in each case. RNAs in dots 1-4 were obtained from female samples, those in dots 5-8 from males. Dots 1, 5, heads; dots 2, 6, thoraces; dots 3, 7, abdomens; dots 4, 8, isolated gonads. The lengths of the two probes and their respective specific activities were similar, thus the intensities of the dots in *a* and *b* represent the relative steady-state levels of the two transcripts.

of a sense/anti-sense hybrid molecule containing both transcripts. However, confirmation of this possibility must await further detailed functional analysis of the 3'-adjacent transcription unit and *in situ* studies of the cell-specific distribution of both species of transcript. While this is only the second known example of overlapping genes in *Drosophila*, such arrangements in eukaryotes may be more common than previously supposed.

We thank Drs Larry Marsh and Ted Wright for sharing unpublished results. A studentship (C.A.S.) and a fellowship (R.D.G.) from the Alberta Heritage Fund for Medical Research are gratefully acknowledged. This research was supported by grant A6377 from NSERC (Canada).

Received 28 March; accepted 28 April 1986.

1. Normark, S. *et al.* *Rev. Genet.* **17**, 499-525 (1983).
2. Henikoff, S., Keene, M. A., Fechtel, K. & Fristrom, J. W. *Cell* **44**, 33-42 (1986).
3. Wright, T. R. F., Hodgetts, R. B. & Seral, A. F. *Genetics* **84**, 267-285 (1976).
4. Pentz, E. S. & Wright, T. R. F. *Genetics* **112**, 843-859 (1986).
5. Eveleth, D. D. & Marsh, J. L. *Genetics* **114** (in the press).
6. Gietz, R. D. & Hodgetts, R. B. *Dev. Biol.* **107**, 142-155 (1985).
7. Spencer, C. A., Gietz, R. D. & Hodgetts, R. B. *Dev. Biol.* **114**, 260-264 (1986).
8. Eveleth, D. D. *et al.* *EMBO J.* (submitted).
9. Birnstiel, M. L., Busslinger, M. & Strub, K. *Cell* **41**, 349-359 (1985).
10. Wickens, M. & Stephenson, P. *Science* **226**, 1045-1051 (1984).
11. Kraminsky, G. P. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **77**, 4175-4179 (1980).
12. Beall, C. J. & Hirsh, J. *Molec. cell. Biol.* **4**, 1669-1674 (1984).
13. Scholnick, S. B., Morgan, B. A. & Hirsch, J. *Cell* **34**, 37-45 (1983).
14. Livingstone, M. S. & Tempel, B. L. *Nature* **303**, 67-70 (1983).
15. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
16. Chen, E. Y. & Seeburg, P. H. *DNA* **4**, 165-170 (1985).
17. Echallier, G. & Ohanessian, A. *C. r. heb. Séanc. Acad. Sci., Paris* **D268**, 1771-1773 (1969).

The RNA required in the first step of chlorophyll biosynthesis is a chloroplast glutamate tRNA

Astrid Schön*, Guido Krupp, Simon Gough†, Sandra Berry-Lowe†, C. Gamini Kannangara† & Dieter Söll‡

Department of Molecular Biophysics and Biochemistry, Yale University, PO Box 6666, New Haven, Connecticut 06511, USA
† Department of Physiology, Carlsberg Laboratory, Gamle Carlsberg Vej 10, 2500 Copenhagen Valby, Denmark

A molecule of chlorophyll is synthesized from eight molecules of δ -aminolevulinic acid (ALA), the universal precursor of porphyrins. The light-regulated conversion of glutamate to δ -aminolevulinic acid in the stroma of greening plastids involves the reduction of glutamate to glutamate-1-semialdehyde and its subsequent transamination¹⁻⁵. The components performing this conversion have been isolated from barley^{1,2} and *Chlamydomonas*⁵ and separated into three fractions by serial affinity chromatography on Blue Sepharose and haem^{1,5} or chlorophyllin-Sepharose². The complete reaction can be performed *in vitro* in a reconstituted assay by combining all three fractions. An RNA is the essential component of the chlorophyllin-Sepharose-bound fraction^{2,3}. By nucleotide sequence analysis, we have now identified this RNA as a chloroplast glutamate acceptor RNA. Glutamate attached by an aminoacyl bond to the 3'-terminal adenosine of this RNA is a substrate for the enzyme(s) which perform the subsequent reactions. This reaction represents a novel role for transfer RNA: participation in the metabolic conversion of its cognate amino acid into another metabolite of low relative molecular mass which subsequently is not used in peptide bond synthesis.

The ultraviolet spectrum of the haem- or chlorophyllin-Sepharose-bound fraction showed the characteristics typical of

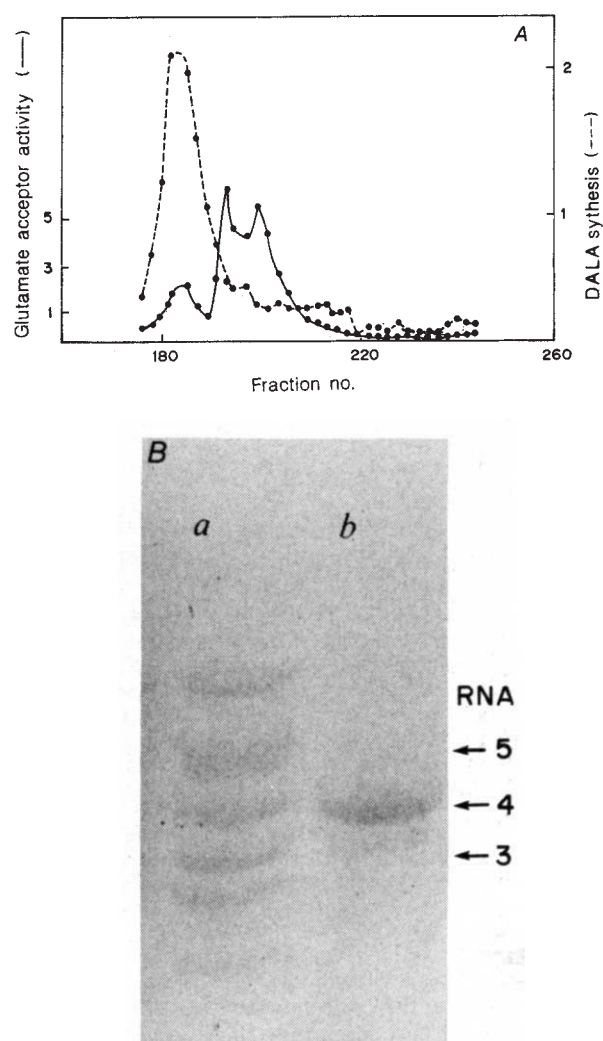


Fig. 1 Purification of RNA^{DALA}. **A**, HPLC fractionation. RNA was recovered from the chlorophyllin-Sepharose-bound fraction by phenol extraction, precipitated with ethanol and injected into a 250×9-mm column packed with triethylammonium chloride-coated 5 μ Hypersil ODS (refs 2, 37). Separation was achieved with a 4-h linear gradient from 0.5–2.3 M ammonium acetate (pH 5.1). Fractions 181–187 were pooled and the RNA was precipitated with ethanol. Glutamate acceptor activity and δ -aminolevulinic acid synthesis were determined as described elsewhere². **B**, Gel purification. For further purification, aliquots of the pooled HPLC fraction (5 μ g dry RNA) were dissolved in 8 M urea containing 0.03% of tracking dyes and loaded into 10-mm-wide slots of 15% polyacrylamide sequencing gels (20×40×0.2 cm); electrophoresis was at 1,500 V until Sigma brilliant blue had reached the bottom of the gel⁸. Gels were stained with toluidine blue 0 and RNA was recovered as described previously³⁸. Lane **a**, RNA species present after HPLC fractionation; lane **b**, separation of partially purified RNAs (recovered from incompletely resolved RNA bands 3–5 in a less extensive electrophoresis). RNAs 1–3 were tested for their ability to function in δ -aminolevulinic acid biosynthesis (see Table 1).

nucleic acids. From these data and the fact that ribonuclease treatment of this fraction destroyed its ability to support δ -aminolevulinic acid formation in the reconstitution assay, it was concluded that RNA is the active principle in the haem- or chlorophyllin-Sepharose-bound material^{2,3}. Phenol extraction of this material and HPLC fractionation led to the separation of three fractions of glutamate acceptor activity (Fig. 1A). Only the first species was able to support δ -aminolevulinic acid formation in the reconstitution assay (Fig. 1A). From this HPLC fraction

* Present address: Institut für Biochemie, Universität Würzburg, D-8700 Würzburg, FRG.

‡ To whom correspondence should be addressed.

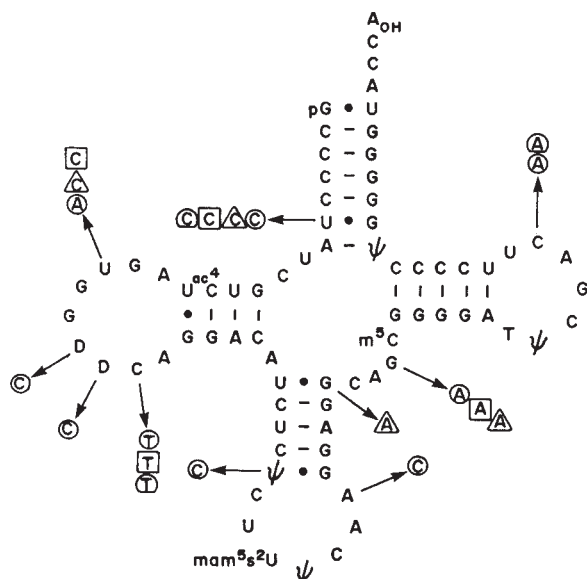


Fig. 2 Nucleotide sequence of RNA^{DALA}. The sequence was determined by base-specific enzymatic and chemical cleavage of the 5'- and 3'-end-labelled RNA⁶⁻⁹. Modified nucleotides were identified by the method of Stanley and Vassilenko⁶. Ambiguities due to the tight secondary structure and the lability of mam^{5s}U were resolved by mobility shift analysis¹¹ of fragments generated either by complete digestion with RNase T₁ or by partial cleavage with aniline⁹. Arrows indicate differences in primary structure from chloroplast tRNA^{Glu} genes of the following species: *Nicotiana tabacum*¹⁹, squashed circle; *Vicia faba*²⁰ and *Pisum sativum*²¹, squares; spinach²², triangle; *Euglena gracilis*²³, round circle. The sequence is identical to a wheat (*Triticum aestivum*) tRNA^{Glu} gene.

a homogeneous active RNA species, referred to as RNA^{DALA}, was purified by polyacrylamide gel electrophoresis (Fig. 1B, Table 1). Analysis of the nucleotide sequence of this RNA was complicated by the fact that it contained a chemically labile base and possessed an extremely stable secondary structure. Chemical and enzymatic sequence analysis of the 3'- and 5'-end-labelled RNA⁶⁻⁹, as well as fingerprinting of post-labelled RNA¹⁰, established the structure of a tRNA (Fig. 2). The labile base in the first position of the anticodon was identified as 5-methylaminomethyl-2-thiouridine (mam^{5s}U). This was established by co-migration of the [5'-³²P]-nucleoside monophosphate¹¹ with the authentic marker nucleotide prepared from *Escherichia coli* tRNA^{Glu}₂ (refs 12, 13) in several thin layer chromatographic systems¹⁴⁻¹⁷ as well as by its sensitivity to cyanogen bromide¹² and its behaviour in a mobility shift analysis¹¹. The nucleotide mam^{5s}U has previously been found exclusively in the tRNAs of prokaryotic organisms¹⁸. Its occurrence in chloroplast tRNA supports the endosymbiont theory of the prokaryotic origin of organelles¹⁹. The anticodon sequence UUC (unmodified) suggests a glutamate tRNA, in agreement with its amino-acid acceptor properties (Fig. 1).

Comparison with known tRNA^{Glu} gene sequences (no other chloroplast tRNA^{Glu} species have been sequenced to date) showed high homology to tRNA^{Glu} genes from chloroplasts of higher plants²⁰⁻²⁵: the known wheat (*Triticum aestivum*) gene has a sequence identical to the barley RNA^{DALA}. The sequence from tobacco (*Nicotiana tabacum*) differs in only three bases, the sequences from broad bean (*Vicia faba*), pea (*Pisum sativum*) and spinach (*Spinacea oleracea*) differ in four positions, whereas the sequence from *Euglena gracilis* shows nine differences (Fig. 2). One distinctive feature of tRNA^{Glu} of chloroplast origin, the A53:A61 pair^{22,23} (the last base pair in the TψC-stem), is also present in this tRNA.

The RNA^{DALA} requires an intact 3'-terminal CCA end for its ability to support δ-aminolevulinate formation. This was demon-

strated by inactivating the glutamate acceptor and δ-aminolevulinate-synthesizing activity of the RNA by partial exonucleolytic digestion with snake venom phosphodiesterase and subsequent restoration by repair of the 3'-terminal CCA end with tRNA nucleotidyl transferase²⁶ (Table 1). The linkage between glutamate and RNA was identified as an authentic aminoacyl bond after labelling the 3'-end of the purified RNA with [α-³²P]ATP by use of nucleotidyl transferase and charging with ¹⁴C-glutamate using the Blue Sepharose-bound enzyme fraction. The terminal ¹⁴C-glutamyl-³²P-adenosine 5'-phosphate was recovered after cleavage with nuclease P₁ and identified by thin layer chromatography¹⁴, using as a marker ¹⁴C-glutamyl-adenosine 5'-phosphate derived from pure aminoacylated *E. coli* tRNA^{Glu}₂ (data not shown). These findings indicate that, as in aminoacyl-tRNA, the α-carboxyl group of glutamate is probably aminoacylated. However, we cannot rule out the possibility that amino-acid activation involves the γ-carboxyl group.

The presence of an enzyme-associated RNA in the chloroplast led to questions of whether, in barley, RNA^{DALA} is encoded in the chloroplast DNA and whether the location of the gene is the same in widely different plant species. Although it is generally believed that organellar RNAs are encoded in the organellar DNA, RNA transport into the mitochondrion has also been described²⁷. Hybridization experiments were performed using genomic and chloroplast DNAs from various plant species, including unicellular algae. To allow comparisons with the available *Hind*III bank of the barley chloroplast genome²⁸, and because the sequence of RNA^{DALA} shows a unique *Hinf*I restriction site in the T-loop region, *Hind*III and *Hinf*I treatments were chosen for Southern hybridizations. To avoid artefacts

Table 1 Characterization of the RNA required for δ-aminolevulinate biosynthesis

Expt	Glutamate to δ-aminolevulinate conversion	
	(c.p.m. above background per 10 µg RNA)	(%)
Expt 1		
No RNA (background)	(19,000)	0
HPLC fraction (Fig. 1a)	200,900	100
Expt 2		
Gel purified RNA (Fig. 1b)		
Band 3	15,900	8
Band 4	243,500	121
Band 5	12,500	6
Expt 3		
Snake venom phosphodiesterase inactivated RNA	3,950	2
RNA repaired with tRNA nucleotidyltransferase	150,700	75

δ-Aminolevulinate synthesis was assayed in a mixture containing 0.5 mM ATP, 1 mM NADPH, 25 mM MgCl₂, 1 mM dithiothreitol (DTT), 0.3 M glycerol, 0.1 M Tricine-NaOH (pH 7.9), 0.25 mg Blue Sepharose-bound protein, 0.4 mg unbound protein and 2.5 µCi ¹⁴C-glutamate (275 Ci mmol⁻¹; 330,000 c.p.m. pmol⁻¹) in a total volume of 0.5 ml. 10–50 µg RNA was used per assay. The conversion of glutamate to δ-aminolevulinate was determined as described elsewhere². For exonuclease inactivation of RNA, 100 µg of HPLC-purified RNA was incubated for 20 min at room temperature with 2 µg snake venom phosphodiesterase in 50 mM Tris-HCl (pH 8), 10 mM MgCl₂. The material was then phenol extracted and ethanol precipitated. Reactivation was performed with 50 µg of this RNA in 0.4 ml, containing 25 µM ATP, 25 µM CTP, 10 mM MgCl₂, 8 mM DTT, 40 mM Tris-HCl (pH 8) and 62.5 µg (0.6 U) of tRNA nucleotidyl transferase. After incubation for 45 min at 37 °C, the RNA was phenol extracted, ethanol precipitated and assayed for reconstitution activity as described above. 3'-End labelling of the gel-purified active RNA species was performed with 5 µg RNA by scaling down these protocols and substituting [5'-³²P]ATP (3,000 Ci mmol⁻¹, 36 pmol) for ATP.

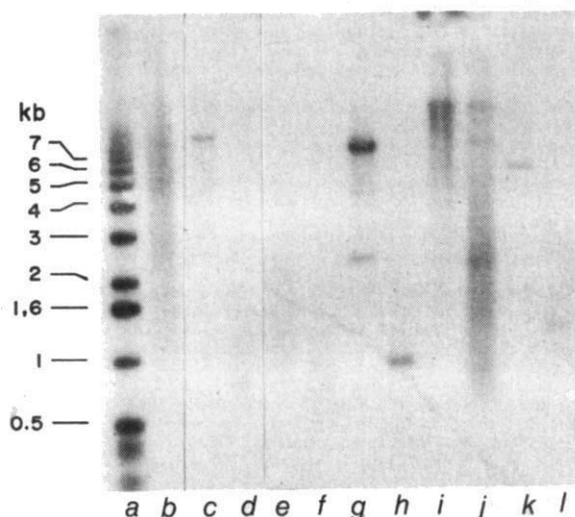


Fig. 3 Hybridization of RNA^{DALA} to genomic and chloroplast DNA. Restricted DNA (1–2 µg) was separated on a 0.7% agarose gel and blotted onto nitrocellulose paper³⁹. Hybridization was performed in 6×SSC, 3×Denhardt's solution and 0.1% sodium pyrophosphate for 1 h at 42 °C. Nonspecifically bound probe was removed by washing the filter with 2×SSC at 42 °C. The lanes contained: a, DNA size markers; b, barley genomic DNA; c, d, barley chloroplast DNA; e, f, tobacco genomic DNA; g, h, tobacco chloroplast DNA; i, j, *Chlamydomonas reinhardtii* genomic DNA; k, l, *Chlamydomonas reinhardtii* chloroplast DNA. All samples were digested with *Hind*III, and d, f, h, j and l were also digested with *Hinf*I.

created by the secondary structure of the RNA, fragments generated by complete digestion with RNase T₁ or by controlled aniline cleavage⁹ (size range 15–30 nucleotides) were used for hybridization. All chloroplast DNAs showed one prominent signal of 7–10 kilobases (kb) in length and two bands of low intensity, whereas the nuclear DNA preparations gave no specific hybridization signal (Fig. 3). *Hinf*I cleaved the prominently hybridizing *Hind*III fragments into DNA of lower relative molecular mass. Although the nuclear genomes obviously contain DNA sequences related to RNA^{DALA}, the presence of a strongly hybridizing DNA fragment in all chloroplast DNAs make it unlikely that this RNA is encoded in the nucleus.

A few examples exist for the involvement of tRNA in reactions other than protein biosynthesis, among them the synthesis of aminoacyl-phosphatidyl-glycerol²⁹, the N-terminal addition of amino acids to proteins³⁰ and the formation of pentapeptide bridges in bacterial cell walls³¹. More recently, tRNA has been reported to be an essential component of the ubiquitin- and ATP-dependent proteolytic system of mammalian cells³². The transamination of glutamyl-tRNA^{Glu} to yield glutamyl-tRNA^{Gln} in *Bacillus megatherium* is an example of enzymatic conversion of an amino acid already linked to a tRNA³³. However, in contrast to the case reported here, glutamine formed in this reaction is then donated into peptide linkage during the

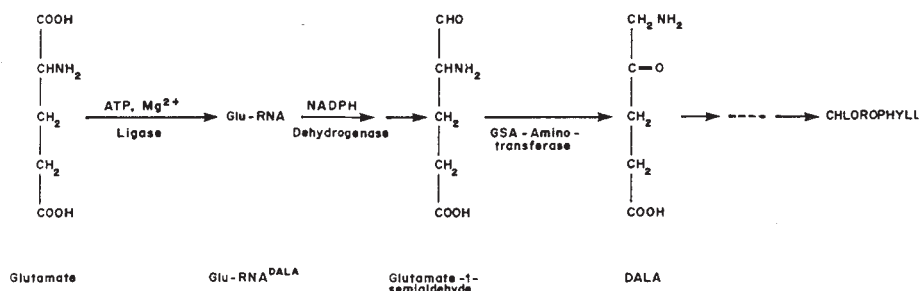
process of protein biosynthesis. For the normal requirements of intermediary metabolism in *Bacillus*, glutamine is formed by glutamine synthase^{33,34}.

Two of the three glutamate isoacceptors in barley chloroplasts are unable to substitute for the specific RNA required in δ-aminolevulinic acid synthesis, although they are efficiently charged by aminoacyl-tRNA synthetases present in the enzyme preparation². This raises the question of whether the RNA^{DALA} is a tRNA species specialized just for δ-aminolevulinic acid formation and not used for protein biosynthesis in the chloroplast. Although the anticodon in RNA^{DALA} is more highly modified than in other known tRNA^{Glu} species, the answer to this question must await the results of *in vitro* protein synthesis experiments comparing all three pure tRNA^{Glu} species. A case of such tRNA specialization is seen in *Staphylococcus*, where some tRNA^{Gly} species participate in cell wall biosynthesis but not in protein synthesis³¹.

The fact that chlorophyll biosynthesis is a light-induced process involving the concerted expression of a number of nuclear and chloroplast genes raises several questions concerning the regulation of gene expression. Of particular interest regarding the occurrence of three chloroplast glutamate isoacceptors is the question of whether they are encoded by a single gene, thus differing only in their base modifications, or whether they are derived from differentially regulated individual genes.

The formation of δ-aminolevulinic acid from glutamate is specific for plant chloroplasts and blue-green algae (for example, cyanobacteria³⁵). An outline of the enzymatic reaction scheme in plants based on our current knowledge is shown in Fig. 4. There are at least three distinct reactions performed by a complex set of enzymes and an RNA moiety. ATP-dependent ligation of glutamate to the RNA^{DALA} by an aminoacyl bond is the reaction initiating this metabolic pathway. It remains to be determined whether this reaction is carried out by the normal chloroplast glutamyl-tRNA synthetase active in the glutamylation of all three chloroplast tRNA species (Fig. 1a) or whether a special enzyme is used. The glutamyl-RNA ligase present in the Blue Sepharose-bound fraction has recently been purified (P. Bruyant, unpublished). The subsequent reduction of the glutamate, activated by attachment to the RNA, is an interesting new type of reaction in which RNA is the 'cofactor' for a dehydrogenase. The fact that only one of the three tRNA^{Glu} isoacceptors is recognized by the dehydrogenase raises interesting questions about the importance for proper recognition of primary structure and of the nucleoside modification pattern of the tRNAs. Depending on the position of the activated carboxyl group in glutamate, an additional enzyme, a hydrolase (indicated by the short arrow in Fig. 4) may be required to cleave glutamate-1-semialdehyde from the RNA^{DALA}. In the final step, free glutamate-1-semialdehyde is then the substrate for a specific aminotransferase. The RNA^{DALA} has been sequenced and identified by Southern hybridization analysis as a chloroplast tRNA^{Glu}. The fact that the RNA hybridized to chloroplast DNA of many different plant species and that a required RNA was also demonstrated for *Chlamydomonas*³ and *Chlorella*³⁶ suggests that the RNA-dependent formation of δ-aminolevulinic acid for chlorophyll synthesis occurs in all plants.

Fig. 4 Scheme of the first steps of chlorophyll synthesis. For details see text.



A.S. was supported by a fellowship of the Deutscher Akademischer Austauschdienst; G.K. was a Postdoctoral Fellow of the Deutsche Forschungsgemeinschaft; S.B.-L. is a Postdoctoral Fellow of the US NSF. We thank Drs H. Cudny and M. Deutscher for their gift of purified tRNA nucleotidyl transferase. Some of the DNA samples from chloroplasts and nuclei were provided by Drs A. Cheung and J. Palmer. We thank Professor Diter von Wettstein for initiating the cooperative work. This work was supported by a NIH grant.

Received 14 February; accepted 15 April 1986.

1. Wang, W.-Y., Gough, S. P. & Kannangara, C. G. *Carlsberg Res. Commun.* **46**, 243-257 (1981).
2. Kannangara, C. G., Gough, S. P., Oliver, R. P. & Rasmussen, S. K. *Carlsberg Res. Commun.* **49**, 417-437 (1984).
3. Huang, D.-D., Wang, W.-Y., Gough, S. P. & Kannangara, C. G. *Science* **225**, 1482-1484 (1984).
4. Kannangara, C. G., Gough, S. P. & von Wettstein, D. in *Development in Plant Biology* Vol. 2 (eds Akoyonoglou, G. & Akoyonoglou, J. H.) 147-160 (Elsevier, Amsterdam, 1978).
5. Wang, W.-Y., Huang, D.-D., Stachon, D., Gough, S. P. & Kannangara, C. G. *Pl. Physiol.* **74**, 569-575 (1984).
6. Stanley, J. & Vassilenko, S. *Nature* **274**, 87-89 (1978).
7. Donis-Keller, H., Maxam, A. M. & Gilbert, W. *Nucleic Acids Res.* **4**, 2527-2537 (1977).
8. Krupp, G. & Gross, H. J. in *The Modified Nucleosides in Transfer RNA II: A Laboratory Manual of Genetic Analysis, Identification and Sequence Determination* (eds Agris, P. F. & Kopper, R. A.) 11-58 (Liss, New York, 1983).
9. Peattie, D. A. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1760-1764 (1979).
10. Domdey, H., Jank, P., Sanger, H. L. & Gross, H. J. *Nucleic Acids Res.* **5**, 1221-1236 (1978).
11. Silberklang, M., Gillum, A. M. & RajBhandary, U. L. *Meth. Enzym.* **59**, 58-109 (1979).

12. Ohashi, Z., Harada, F. & Nishimura, S. *FEBS Lett.* **20**, 239-241 (1972).
13. Munnings, K. O. & Chang, S. H. *Biochem. biophys. Res. Commun.* **46**, 1837-1842 (1972).
14. Nishimura, S. in *tRNA: Structure, Properties and Recognition* (eds Schimmel, P. R., Soll, D. & Abelson, J. N.) 551-552 (Cold Spring Harbor Laboratory, New York, 1979).
15. Saneyoshi, M. & Nishimura, S. *Biochim. biophys. Acta* **204**, 389-399 (1979).
16. Gupta, R., Randerath, E. & Randerath, K. *Nucleic Acids Res.* **3**, 2915-2921 (1976).
17. Raba, M. *et al. Eur. J. Biochem.* **97**, 305-318 (1979).
18. Sprinzl, M., Moll, J., Meissner, F. & Hartmann, T. *Nucleic Acids Res.* **13**, r1-r49 (1985).
19. Fox, G. E. *et al. Science* **209**, 457-463 (1980).
20. Quigley, F. & Weil, J. H. *Curr. Genet.* **9**, 495-503 (1985).
21. Ohme, M., Kamogashira, T., Shinozaki, K. & Sugiura, M. *Nucleic Acids Res.* **13**, 1045-1056 (1985).
22. Kuntz, M., Weill, J. H. & Steinmetz, A. *Nucleic Acids Res.* **12**, 5037-5047 (1984).
23. Rasmussen, O. F., Stummann, B. M. & Henningsen, K. W. *Nucleic Acids Res.* **12**, 9143-9153 (1984).
24. Holschuh, K., Bottomley, W. & Whitfield, P. R. *Pl. molec. Biol.* **3**, 313-317 (1984).
25. Hollingsworth, M. J. & Hallick, R. B. *J. biol. Chem.* **257**, 12795-12799 (1982).
26. Silberklang, M., Gillum, A. M. & RajBhandary, U. L. *Nucleic Acids Res.* **4**, 4091-4108 (1978).
27. Martin, R. P., Schneller, J. M., Stahl, A. J. C. & Dirheimer, G. *Nucleic Acids Res.* **4**, 3497-3511 (1977).
28. Poulsen, C. *Carlsberg Res. Commun.* **48**, 57-80 (1983).
29. Nesbitt, J. A. & Lennartz, W. J. *J. biol. Chem.* **243**, 3088-3095 (1968).
30. Leibowitz, M. J. & Soffer, R. L. *Biochem. biophys. Res. Commun.* **36**, 47-53 (1969).
31. Bumsted, R. M., Dahl, J. M., Soll, D. & Strominger, J. L. *J. biol. Chem.* **249**, 4787-4796 (1974).
32. Ciechanover, A., Wolin, S., Steitz, J. A. & Lodish, H. F. *Proc. natn. Acad. Sci. U.S.A.* **82**, 1341-1345 (1985).
33. Wilcox, M. & Nirenberg, M., *Proc. natn. Acad. Sci. U.S.A.* **61**, 229-236 (1968).
34. Henner, D. J. & Hoch, J. A. *Microbiol. Rev.* **44**, 57-82 (1980).
35. McKie, J., Lucas, C. & Smith, A. *Phytochemistry* **20**, 1547-1549 (1981).
36. Weinstein, J. D. & Beale, S. I. *Archs Biochem. Biophys.* **239**, 87-93 (1985).
37. Bischoff, R. E., Graeser, E. & McLaughlin, L. J. *J. Chromat.* **257**, 305-315 (1983).
38. Beier, H., Barciszewska, M., Krupp, G., Mitnacht, R. & Gross, H. J. *EMBO J.* **3**, 351-356 (1984).
39. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning, a Laboratory Manual* (Cold Spring Harbor Laboratories, New York, 1982).

Quantitative analysis of structure-activity relationships in engineered proteins by linear free-energy relationships

Alan R. Fersht, Robin J. Leatherbarrow & Tim N. C. Wells

Department of Chemistry, Imperial College of Science and Technology, London SW7 2AY, UK

Protein engineering is being used increasingly to study the fine details of the structure and activity of enzymes. How can small effects of mutation on activity be reliably quantified and systematized, and artefacts be recognized? A traditional means of doing this in organic chemistry is the use of linear free-energy relationships that link changes in rate constant for a reaction to changes in its equilibrium constant as the structure of the reagents is altered—Bronsted or Hammett plots¹. We now find that the same type of plot may be applied to enzymatic reactions for variation of the structure of an enzyme with mutation. The activities of many mutant tyrosyl-transfer RNA synthetases fit structure-activity relationships which relate the rate constant for the formation of enzyme-bound tyrosyl adenylate (E.Tyr-AMP) to its equilibrium constant with enzyme-bound tyrosine and ATP (E.Tyr.ATP). This reaction results in an increase in binding energy between certain side chains of the enzyme and the side chain of tyrosine and the ribose ring of ATP. Plots of rate against equilibrium constant for a series of enzymes mutated in the relevant positions indicate that 71% of the binding energy change occurs on formation of the transition state for the chemical reaction and 90% occurs on formation of the E.Tyr-AMP.PP_i complex. Other mutations show a different behaviour which is diagnostic of residues that specifically bind the transition state. Linear free-energy plots show trends and allow exceptions to be readily noted. That the activities of a large number of mutants conform to linear free-energy equations is the best evidence yet that mutation of an enzyme can probe general properties and trends in the relationship between structure and activity.

Mutants of the tyrosyl-tRNA synthetase, which catalyses the formation of tyrosyl adenylate [equation (1)], have been gener-

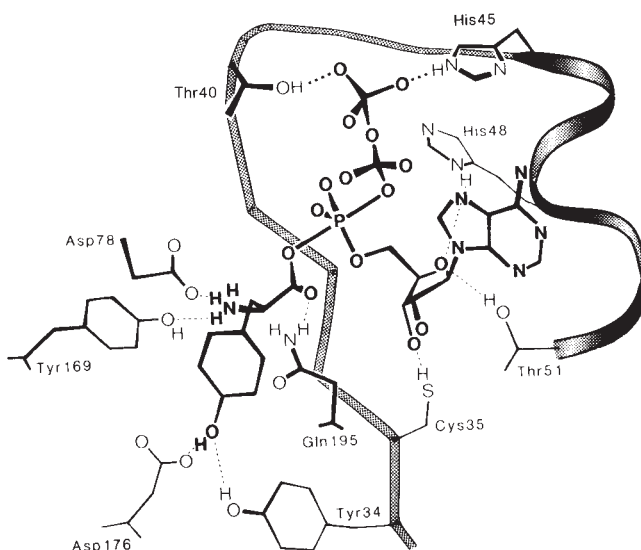


Fig. 1 Side chains of the tyrosyl-tRNA synthetase from *Bacillus stearothermophilus* which form hydrogen bonds with the transition state of ATP and tyrosine during the formation of tyrosyl adenylate (from ref. 5).

ated by site-directed mutagenesis and studied by kinetics²⁻⁴.



These mutants have been altered in side chains that make hydrogen bonds with the substrates (Fig. 1). Model building followed by mutagenesis has revealed that there is a major catalytic factor caused by just two side chains that contribute binding energy to the γ-phosphate of ATP only when it is in the transition state (Thr 40 and His 45)⁵. Further, the binding energies of groups far removed from the site of reaction are also used to increase the chemical rate constants for the reaction (Tyr 34, Cys 35, His 48, Thr 51 and Tyr 169)⁶. Subsequent studies have measured the complete free-energy profiles for activation of tyrosine by the mutant enzymes^{7,8}. The comparison of these

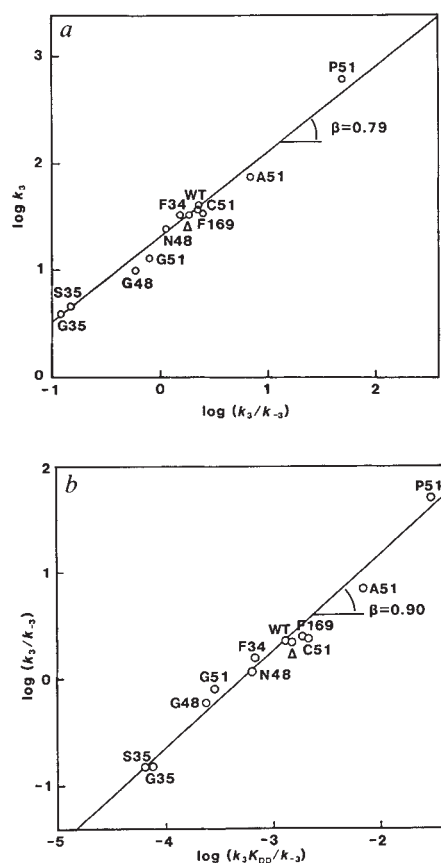


Fig. 2 *a*, Plot of $\log k_3$ against $\log (k_3/k_{-3})$ for equation (5) for mutation of Tyr 34, Cys 35, His 48, Thr 51 and Tyr 169. The mutations are indicated by the one-letter code apart from Δ , which represents the truncated enzyme lacking the tRNA-binding domain¹⁰. Data are from refs 6 and 7 and T.N.C.W. (unpublished). The plot is equivalent to a plot of $G_{[E.Tyr.ATP]} - G_{E.Tyr.ATP}$ against $G_{E.Tyr.AMP.PP_i} - G_{E.Tyr.ATP}$, where G is Gibbs free energy and $[E.Tyr.ATP]$ is the transition state. *b*, Plot of $\log (k_3/k_{-3})$ against $\log (k_3K_{pp}/k_{-3})$. This is equivalent to a plot of $G_{E.Tyr.AMP.PP_i} - G_{E.Tyr.ATP}$ against $G_{E.Tyr.AMP} - G_{E.Tyr.ATP}$. Note that individual variations in rate and equilibrium constants are seen as deviations from the regression line.

with the profile for wild-type enzyme gives the apparent binding energy of the relevant side chain with the substrates, transition states and intermediates throughout the reaction. These data have been analysed in each individual case to show how the changes in binding energy affect the rate and equilibrium constants for the interconversion of free and enzyme-bound species. Several of the side chains that bind the side chain of tyrosine and the ribose ring of ATP appear to stabilize the enzyme-bound tyrosyl adenylate complex (E.Tyr-AMP) more than the transition state for its formation and more still than the initial enzyme-substrate ternary complex (E.Tyr.ATP)^{7,8}. By this type of analysis, we are building up a picture of how each side chain contributes to catalysis.

The question most frequently asked about this approach is: How do we know that alteration of a side chain by mutagenesis does not cause additional structural changes in the enzyme which lead to artefactual results? X-ray crystallography of mutants can provide important evidence but may miss a series of small changes or the existence of conformational changes in solution. Further, although crystallography and spectroscopy can solve the structures of the stable states of the enzyme and its substrates, they do not give direct information on the structure

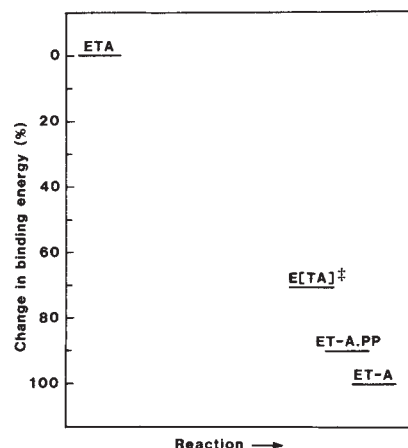


Fig. 3 Summary of the average binding-energy changes calculated for residues Tyr 34, Cys 35, His 48, Thr 51 and Tyr 169 on going from the E.Tyr.ATP complex (ETA) to the transition state ($E[TA]^\ddagger$) to the first intermediate, E.Tyr-AMP.PP_i (ET-A.PP), to the final intermediate, E.Tyr-AMP (ET-A).

which is important for the determination of rate, the transition state. The only way this can be probed experimentally at present is by kinetics. A possible way of both probing the structure of the transition state and detecting artefacts is the use of linear free-energy plots. In these, plots of rate constant against equilibrium constant are made for a reaction in which the structures of the reagents are systematically varied. The plots can give information on how much the transition state resembles starting materials or products. The plots assume that there is the relationship between the equilibrium constant K of a reaction and the rate constant k for the formation of products of the general form:

$$k = AK^\beta \quad (2)$$

where A and β are constants. The plot of $\log k$ against $\log K$ when equation (2) holds is a straight line of slope β [equation (3)].

$$\log k = \text{constant} + \beta \log K \quad (3)$$

Since $\log k$ is proportional to the free energy of activation of the reaction, $\Delta G^\ddagger$, and $\log K$ is proportional to the free-energy change for the equilibrium, ΔG_e , equation (3) is equivalent to:

$$\Delta G^\ddagger = \text{constant} + \beta \Delta G_e \quad (4)$$

A qualitative interpretation of equations (2) to (4) is that, all things being equal, a value of β which is close to zero means that the transition state resembles starting materials, and a value of β close to 1 means that the transition state resembles products. Linear free-energy relationships in simple organic reactions generally measure the inductive effects of groups on rate and equilibrium constants. In enzymatic reactions, however, changes in structure remote from the seat of reaction result in changes of binding energy (ref. 9 and A.R.F., unpublished). Can linear free-energy relationships be applied to the changes in rate and equilibrium constants which occur when enzyme structure is varied as in a site-directed mutagenesis experiment? If so, the values of β will, in general, measure the fractions of total changes in binding energy⁹.

We have measured all the necessary rate and equilibrium constants defined in equation (5) to construct plots for different stages of the reaction.



As Fig. 2 shows, a plot of $\log k_3$ against $\log (k_3/k_{-3})$ for mutants in the side chains which bind the side chain of tyrosine and

the ribose ring of ATP fits an excellent straight line of slope $\beta = 0.79$. This means that 79% of the binding-energy change on E.Tyr.ATP going to E.Tyr-AMP.PP_i is realized in the transition state for the reaction. Also plotted in Fig. 2 is $\log(k_3/k_{-3})$ against $\log(k_3K_{pp}/k_{-3})$. This is in reality a plot of the change in binding energy on going from E.Tyr.ATP to E.Tyr-AMP.PP_i against the change in going from E.Tyr.ATP to E.Tyr-AMP + PP_i. The slope of 0.9 shows that 90% of the change in binding energy on forming E.Tyr-AMP occurs at the stage of formation of E.Tyr-AMP.PP_i. Figure 3 summarizes the changes in binding energy. There is a gain in binding energy on going from the ternary complex E.Tyr.ATP to the enzyme-bound tyrosyl adenylate complex, E.Tyr-AMP; 71% of this binding energy is realized in the transition state and 90% in the other intermediate, E.Tyr-AMP.PP_i.

The above type of behaviour applies when the residues concerned show a uniformly progressive change in binding energy as the reaction proceeds. A different pattern is expected when the mutation involves residues that bind the transition state of the substrates well but bind the unreacted substrates and products only poorly. Such behaviour should lead to values for β which are much greater than 1. For example, in the extreme case, where a residue binds only at the stage of the transition state, mutation of that residue will not affect the equilibrium constant of the reaction but will affect the rate; that is, β should tend to infinity. This is found quite dramatically with residues Thr 40 and His 45 (Fig. 4), which have already been implicated

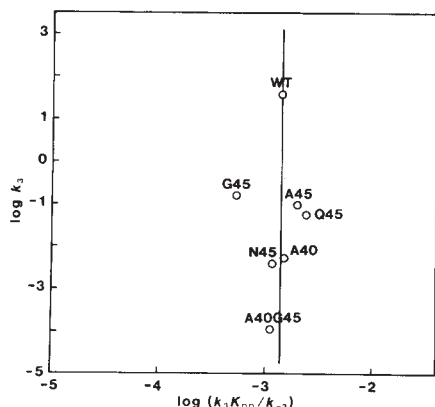


Fig. 4 Plot of $\log k_3$ against $\log(k_3K_{pp}/k_{-3})$ for the mutation of residues Thr 40 and His 45 which are involved in transition-state binding only⁵. This is equivalent to a plot of $G_{[E.Tyr.ATP]} - G_{E.Tyr.ATP}$ against $G_{E.Tyr-AMP} - G_{E.Tyr.ATP}$. Data are from ref. 5 and R.J.L. (unpublished).

in such a mode of catalysis⁵. The very high value of β found indicates that they are highly specific for the transition state.

It is quite remarkable that plots in Fig. 2 are so similar to those found for linear free-energy relationships in simple organic reactions. Indeed, the fit to such plots is as good as many examples from physical organic chemistry. Why are the linear free-energy relationships followed so well considering the likelihood of disturbing the structure of the protein? The answer is probably because we have restricted our mutations to the most conservative changes, which have made only small perturbations in the structure.

The demonstration that certain structural changes of the tyrosyl-tRNA synthetase cause changes in rate which may be described satisfactorily by linear free-energy equations is important for the following reasons. First, the equations combine the data from many experiments and so bring order to them. Second, they quantify aspects of the transition state structure; in particular, they show how much the interactions in the transition

state resemble those in the enzyme-substrate and the enzyme-intermediate complexes. Finally, the equations show trends; any mutant that deviates from the trend is immediately apparent.

The last point is particularly important. The results of any one individual mutation can be challenged on the grounds that the mutation might cause a significant change in the protein structure. However, when a whole series of mutational experiments can be fitted to a linear free-energy equation, there is overwhelming evidence that each mutant shows part of a general phenomenon of the relationship between structure and activity. This type of analysis can be applied to any enzymatic reaction where both rate and equilibrium constants may be measured.

Received 6 May; accepted 5 June 1986.

1. Jencks, W. P. *Chem. Rev.* **85**, 511-517 (1985).
2. Winter, G., Fersht, A. R., Wilkinson, A. J., Zoller, M. & Smith, M. *Nature* **299**, 756-758 (1982).
3. Fersht, A. R. *et al. Angew. Chem. int. Edn Engl.* **23**, 467-473 (1984).
4. Fersht, A. R. *et al. Nature* **314**, 235-238 (1985).
5. Leatherbarrow, R. J., Winter, G. & Fersht, A. R. *Proc. natn. Acad. Sci. U.S.A.* **82**, 7840-7844 (1985).
6. Wells, T. N. C. & Fersht, A. R. *Nature* **316**, 656-657 (1985).
7. Wells, T. N. C. & Fersht, A. R. *Biochemistry* **25**, 1881-1886 (1986).
8. Ho, C. K. & Fersht, A. R. *Biochemistry* **25**, 1891-1897 (1986).
9. Fersht, A. R. *Enzyme Structure and Mechanism* 2nd edn, Ch. 12 (Freeman, San Francisco, 1985).
10. Waye, M. M. Y., Winter, G., Wilkinson, A. J. & Fersht, A. R. *EMBO J.* **2**, 1827-1829 (1983).

Directional electron transfer in ruthenium-modified horse heart cytochrome *c*

Rolf Bechtold, Christa Kuehn, Chris Lepre & Stephan S. Isied*

Department of Chemistry, Rutgers, State University of New Jersey, New Brunswick, New Jersey 08903, USA

Cytochrome *c* can be modified by $[(NH_3)_5Ru^{II/III}]$ specifically at the imidazole moiety of histidine 33, and we have recently discussed the thermodynamics and kinetics of electron transfer within this modified protein¹⁻⁵. X-ray crystal structures of the oxidized and reduced forms of tuna cytochrome *c*⁶ indicate that the separation between the haem group of cytochrome *c* and the ruthenium label is 12-16 Å. Internal electron transfer from the $[(NH_3)_5Ru^{II}]$ centre to the Fe(III) haem centre occurs with a rate constant $k \approx 53 s^{-1}$ (25 °C) ($\Delta H^\ddagger = 3.5 kcal mol^{-1}$, $\Delta S^\ddagger = -39 EU$), as measured by pulse radiolysis. The measured unimolecular rate constant¹, $k \approx 53 s^{-1}$, is on the same timescale as a number of conformational changes that occur within the cytochrome *c* molecule⁷⁻⁹. These results raise the question of whether electron transfer or protein conformational change is the rate limiting step in this process. We describe here an experiment that probes this intramolecular electron transfer step further. It involves reversing the direction of electron transfer by changing the redox potential of the ruthenium label. Electron transfer in the new ruthenium-cytochrome *c* derivative described here is from haem(II) to the Ru(III) label, whereas in $(NH_3)_5Ru$ -cytochrome *c* the electron transfer is from Ru(II) to haem(III). Intramolecular electron transfer from haem(II) to Ru(III) in the new ruthenium-cytochrome *c* described here proceeds much slower ($> 10^5$ times) than the electron transfer from Ru(II) to haem(III) in the $(NH_3)_5Ru$ -cytochrome *c*. We therefore conclude that electron transfer in cytochrome *c* is directional, with the protein envelope presumably involved in this directionality.

* To whom correspondence should be addressed.

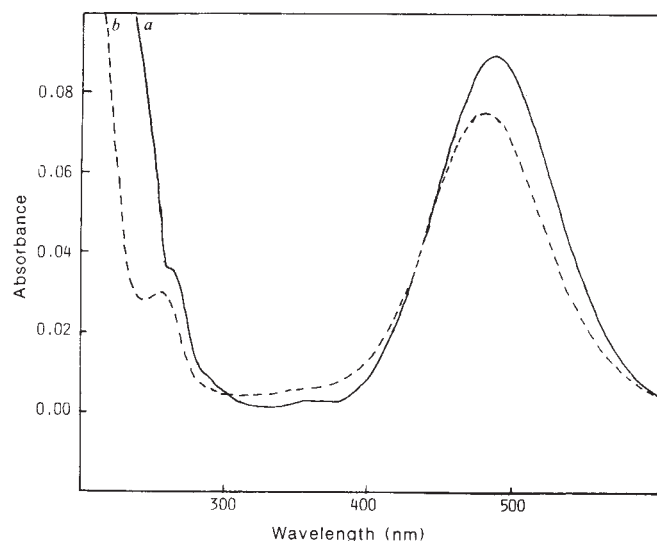


Fig. 1 *a*, Ultraviolet-visible spectrum of the $\text{Ru}^{\text{II}}(\text{NH}_3)_4(\text{isn})$ -peptide fragment containing His 33. *b*, Ultraviolet-visible spectrum of the model complex, $\text{trans}-[\text{Ru}(\text{NH}_3)_4(\text{isn})\text{-N-acetylhistamine}]^{2+}$.

The new ruthenium-modified cytochrome *c* has *trans*- $[\text{Ru}(\text{NH}_3)_4(\text{N} \begin{array}{c} \diagup \text{C}=\text{O} \\ \diagdown \text{NH}_2 \end{array} \text{C}_6\text{H}_4)]$ covalently bound to His 33

and was synthesized using techniques described earlier³. Ruthenium analysis by direct coupled plasma atomic absorption spectroscopy showed the presence of 1 mol of ruthenium per mol of iron. The ultraviolet-visible spectra of the oxidized derivative indicated no difference from that of native cytochrome *c*(III). Comparison between the tryptic digests of the native cytochrome *c* and the $[\text{Ru}(\text{NH}_3)_4(\text{isn})\text{-cyt } c]$ (isn = isonicotinamide) by HPLC showed the presence of a new fragment in the ruthenium-modified product. The amino-acid composition of this fraction (Asp, 1.04; Thr, 0.97; Gly, 2.01; Leu, 1.62; His, 0.94; Pro, not detected) corresponds to amino acids 28–35 of cytochrome *c*. (The new cleavage site for trypsin may have resulted from the hydrophobicity of the isonicotinamide ligand in combination with the positive charge of the ruthenium centre.)

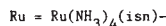
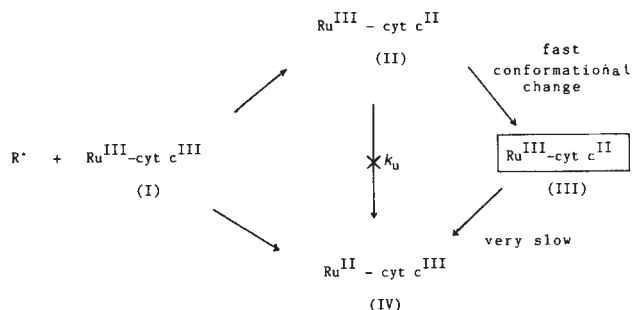
The ultraviolet-visible spectrum of this peptide fragment after reduction with ascorbic acid (Fig. 1*a*) closely resembles that of the model compound $\text{trans}-[\text{Ru}^{\text{II}}(\text{NH}_3)_4(\text{isn})(\text{N-acetylhistamine})]$ (Fig. 1*b*). Figure 2*a* shows a differential pulse polarogram (DPP) of $[\text{Ru}(\text{NH}_3)_4(\text{isn})\text{-cyt } c]$. Two waves are observed, one corresponding to the haem iron (II/III) of cytochrome *c* (0.26 V versus normal hydrogen electrode (NHE)) and the other corresponding to $[\text{Ru}^{\text{II/III}}(\text{NH}_3)_4(\text{isn})]$ (0.44 V versus NHE). Figure 2*b* shows the DPP of $[(\text{Ru}(\text{NH}_3)_5\text{-cyt } c)]$ for comparison¹.

A spectrophotometric titration of the fully reduced ruthenium-modified cytochrome *c*, $[\text{Ru}^{\text{II}}(\text{NH}_3)_4(\text{isn})\text{-cyt } c^{\text{II}}]$, with the $\text{trans}-[\text{Ru}(\text{NH}_3)_4(\text{isn})_2]^{3+}$ oxidant ($E^\circ = 0.70$ V; estimated from ref. 10), was carried out to determine the stoichiometry of the oxidation. Colour changes in the $\text{trans}-[\text{Ru}(\text{NH}_3)_4(\text{isn})_2]^{3+}$ species during titration enable the oxidation of the haem and the ruthenium sites to be observed sequentially. Two mol of $\text{trans}-[\text{Ru}(\text{NH}_3)_4(\text{isn})_2]^{3+}$ oxidant were required per mol of the cytochrome *c* derivative.

The kinetics of intermolecular oxidation of the haem site in $[\text{Ru}^{\text{II}}(\text{NH}_3)_4(\text{isn})\text{-cyt } c^{\text{II}}]$ with $[\text{Fe}(\text{CN})_6]^{3-}$ or $\text{trans}-[\text{Ru}(\text{NH}_3)_4(\text{isn})_2]^{3+}$ is similar to the oxidation of native cytochrome *c*(II) by inorganic oxidizing agents¹¹. This indicates similar environments of the haem site in the native and the modified protein.

As Fig. 2 shows, the two ruthenium-cytochrome *c* derivatives, $[\text{Ru}(\text{NH}_3)_4(\text{isn})\text{-cyt } c]$ and $[\text{Ru}(\text{NH}_3)_5\text{-cyt } c]$, possess $\text{Ru}(\text{II/III})$ potentials approximately equidistant in the positive and negative directions, respectively, from the potential of $\text{Fe}(\text{II/III})$ haem ($\Delta E^\circ = 0.18$ V and 0.11 V). Thus, on the basis of driving force¹², the rate of intramolecular electron transfer from cytochrome *c*(II) to $[\text{Ru}^{\text{III}}(\text{NH}_3)_4(\text{isn})]$ should be equal to or slightly greater than the rate of intramolecular electron transfer measured from $[\text{Ru}^{\text{II}}(\text{NH}_3)_5]$ to cytochrome *c*(III) ($k \approx 53 \text{ s}^{-1}$).

Reduction of a solution of $[\text{Ru}^{\text{III}}(\text{NH}_3)_4(\text{isn})\text{-cyt } c^{\text{III}}]$ in 0.1 M phosphate buffer pH 7 (scheme 1) with the radicals derived from



Scheme 1

either isopropanol or formate ($\text{R}^\bullet$) using pulse radiolysis techniques is shown in Fig. 3.

The haem site is reduced at the expected rate¹, but no intramolecular electron transfer is observed from the $\text{Fe}(\text{II})$ haem to the $[\text{Ru}^{\text{III}}(\text{NH}_3)_4(\text{isn})]$ site on the timescale of the experiment (2 s). This result is completely unexpected when compared with the $[\text{Ru}(\text{NH}_3)_5\text{-cyt } c]$ derivative. In both the $[\text{Ru}(\text{NH}_3)_4(\text{isn})\text{-cyt } c]$ and the $[\text{Ru}(\text{NH}_3)_5\text{-cyt } c]$ derivatives, the ruthenium is bound to the identical His 33 site and the difference in driving

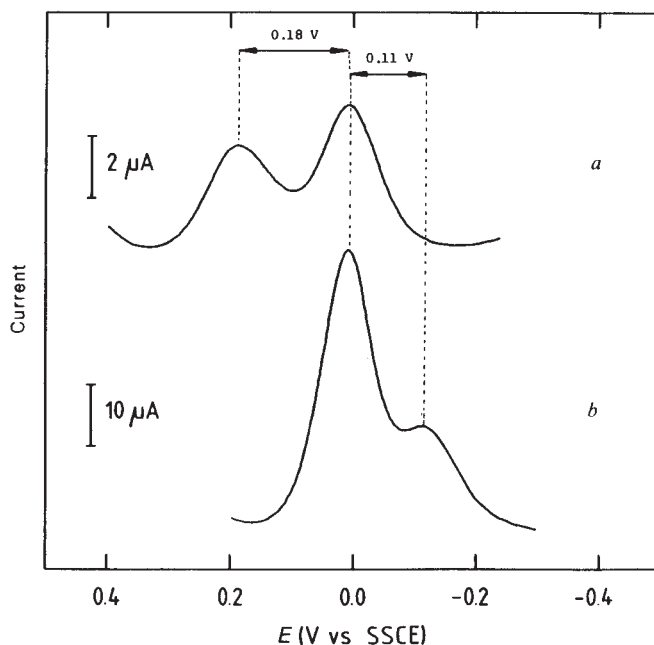


Fig. 2 Differential pulse polarograms of $[\text{Ru}(\text{NH}_3)_4(\text{isn})\text{-cyt } c]$ (*a*) and $[\text{Ru}(\text{NH}_3)_5\text{-cyt } c]$ (*b*); volts versus standard saturated calomel electrode (SSCE); 2-mm gold disk electrode; scan rate = 2 mV s^{-1} ; pulse amplitude, 25 mV; approximately $0.25 \mu\text{mol } [\text{Ru}^{\text{III}}\text{-cyt } c^{\text{III}}]$ in 0.1 M NaClO_4 , 0.01 M bipyridine, 0.08 M phosphate buffer, pH 7.

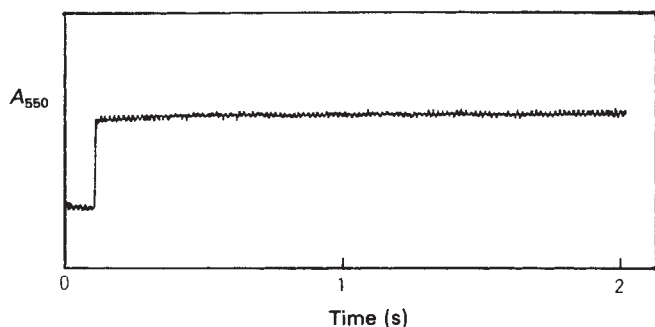


Fig. 3 Absorption change at 550 nm for the reduction of $[\text{Ru}^{\text{III}}(\text{NH}_3)_4(\text{isn})\text{-cyt } c^{\text{III}}]$ with isopropanol radical. The small absorbance increase at long times in the pulse radiolysis experiment is due to the reduction of the haem site in $\text{Ru}(\text{III})\text{-cyt } c(\text{III})$ by secondary radicals. A similar increase in absorbance is observed at long times with the native cytochrome $c(\text{III})$.

force between each ruthenium complex and the haem iron is of a similar magnitude.

Scheme 1 shows the proposed sequence of events after reduction of $[\text{Ru}^{\text{III}}(\text{NH}_3)_4\text{isn-cyt } c^{\text{III}}]$ using pulse radiolysis or chemical techniques. Following reduction, species II and IV are rapidly formed. k_u is the expected rate constant for intramolecular electron transfer between species II (the kinetically stable product) and species IV (the thermodynamically stable product). However, no such reaction is observed and

Table 1 Chemical and pulse radiolysis reduction of $[\text{Ru}^{\text{III}}(\text{NH}_3)_4\text{isn-cyt } c^{\text{III}}]$

Reductant*	% Haem reduced	% $[\text{Ru}^{\text{III}}(\text{NH}_3)_4(\text{isn})\text{-His } 33]$ reduced
$[\text{Ru}(\text{NH}_3)_6]^{2+}$	80 ± 10	20 ± 10
CO_2^-	40 ± 5	60 ± 5
$(\text{CH}_3)_2\text{COH}$	60 ± 5	40 ± 5

Data were obtained by monitoring the increase in absorbance at wavelength $\lambda = 550$ nm for the reduction of the haem site and the increase in absorbance at $\lambda = 504$ nm for the reduction of the $\text{Ru}(\text{III})$ site.

* One mol of $[\text{Ru}(\text{NH}_3)_6]^{2+}$ was used per mol $[\text{Ru}^{\text{III}}(\text{NH}_3)_4\text{isn-cyt } c^{\text{III}}]$. CO_2^- and $(\text{CH}_3)_2\text{COH}$ were formed by pulse radiolysis from formate and 2-propanol, respectively.

instead species II undergoes a rapid (conformational) change to species III. Species III is long-lived (several hours), but can be converted to species IV by the use of a mediator, $[\text{Ru}(\text{NH}_3)_5\text{pyridine}]^{3+}$ (redox potential, $E^\circ = 0.3$ V versus NHE, in between the reduction potentials of cytochrome c and the $[\text{Ru}^{\text{III}}(\text{NH}_3)_4\text{isn}]$ sites). This mediator oxidizes the haem $c(\text{II})$, resulting in the formation of $[\text{Ru}(\text{NH}_3)_5\text{pyridine}]^{2+}$ which in turn reduces the $[\text{Ru}^{\text{III}}(\text{NH}_3)_4\text{isn}]$ site on cytochrome c by intermolecular reactions, as observed from spectral changes on the haem and the ruthenium sites. If species III is generated and kept without this mediator, it decays by a bimolecular pathway on a timescale of several hours ($k = 10 \pm 8 \text{ M}^{-1} \text{ s}^{-1}$; studied over a concentration range of 5–50 μM). The chemical reduction of $[\text{Ru}^{\text{III}}(\text{NH}_3)_4\text{isn-cyt } c^{\text{III}}]$ with 1 mol of $[\text{Ru}(\text{NH}_3)_6]^{2+}$ resulted in the formation of the kinetic intermediates II and III (scheme 1 and Table 1), as observed in the pulse radiolysis experiment.

The combined results of these ruthenium modification experiments constitute strong evidence for directional electron transfer in horse heart cytochrome c , where the intramolecular electron transfer pathway for the reduction of the haem site is more facile than that for the oxidation of the haem by the $\text{Ru}(\text{III})$ complex. This type of behaviour probably results from a conformational change in the protein, such that when cytochrome $c(\text{III})$ is reduced to cytochrome $c(\text{II})$, the reduced haem possesses a high kinetic barrier for intramolecular oxidation from a long distance (12–16 Å). It is noteworthy that intermolecular oxidation of cytochrome $c(\text{II})$ by redox reagents such as cobalt phenanthrolines and ferrocene is very rapid, presumably because there is no restriction on the approach of the oxidant to a favourable haem site for reduction, an option not available in this long-range intramolecular oxidation reaction in $\text{Ru}(\text{NH}_3)_4\text{isn-cyt } c$.

Earlier work on cytochrome c by Tabushi *et al.*^{13,14} using stopped-flow circular dichroism techniques showed the presence of intermediates formed during the reduction of cytochrome $c(\text{III})$, but not during re-oxidation. These results led the authors to suggest that the reduction pathway is different from the oxidation pathway¹³. Furthermore, theoretical and experimental evidence showed conformational differences between the oxidized and reduced forms of cytochrome c in solution^{15–18}.

This directionality of electron transfer in cytochrome c may have great significance in the biological function of cytochrome c , as it may mean that cytochrome c and perhaps other electron transfer proteins undergo electron transfer accompanied by conformational changes. These conformational changes can modify the barrier for electron transfer in different directions. Our observations also call for a differentiation between the formalisms used to describe electron transfer in small molecules and those needed for electron transfer proteins. For example, the meaning of the term 'self exchange' for an electron transfer protein should be redefined, since a multi-step process is involved in the electron exchange of these proteins.

We thank Dr Harold Schwarz (Brookhaven National Laboratories) for his help in the pulse radiolysis experiments, and acknowledge the support of NIH grant GM 26324 and NSF grant 840552. S.S.I. is the recipient of NIH Career Development Award (AM 00732) (1980–85) and a Camille and Henry Dreyfus Teacher Scholar Award (1981–85).

Received 24 March; accepted 18 April 1986.

- Isied, S. S., Kuehn, C. & Worosila, G. *J. Am. chem. Soc.* **106**, 1722–1726 (1984).
- Isied, S. S., Worosila, G. & Atherton, S. J. *J. Am. chem. Soc.* **104**, 7659–7661 (1982).
- Yocum, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **91**, 7052–7055 (1982).
- Nocera, D. G., Winkler, J. R., Yocum, K. M., Bordignon, E. & Gray, H. B. *J. Am. chem. Soc.* **106**, 5145–5150 (1984).
- Yocum, K., Winkler, J. R., Nocera, D. G., Bordignon, E. & Gray, H. B. *Chemica Scr.* **21**, 29–33 (1983).
- Takano, T. & Dickerson, R. E. *J. molec. Biol.* **153**, 95–115 (1981).
- Creutz, C. & Sutin, N. *Proc. natn. Acad. Sci. U.S.A.* **70**, 1701–1703 (1973).
- Creutz, C. & Sutin, N. *J. biol. Chem.* **249**, 6788–6795 (1974).
- Moore, G. *et al. Discuss. Faraday Soc.* **74**, 311–379 (1982).
- Zwickel, A. M. & Creutz, C. *Inorg. Chem.* **10**, 2395–2399 (1971).
- Wherland, S. & Gray, H. B. in *Biological Aspects of Inorganic Chemistry* (eds Addison, A. W., Cullen, W. R., Dolphin, D. & James, B. R.) 289–368 (Wiley, New York, 1977).
- Isied, S. S. & Taube, H. *Inorg. Chem.* **15**, 3070–3075 (1976).
- Tabushi, I., Yamamura, K. & Nishiyama, T. *J. Am. chem. Soc.* **101**, 2785–2787 (1979).
- Tabushi, I., Yamamura, K. & Nishiyama, T. *Tetrahedron Lett.* **49**, 4921–4924 (1978).
- Bosshard, H. R. & Zürrer, M. *J. biol. Chem.* **255**, 6694–6699 (1980).
- Rackovsky, S. & Goldstein, D. A. *Proc. natn. Acad. Sci. U.S.A.* **81**, 5901–5905 (1984).
- Williams, G., Clayden, N. J., Moore, G. R. & Williams, R. J. P. *J. molec. Biol.* **183**, 447–460 (1985).
- Margoliash, E. & Schejter, A. *Adv. Protein Chem.* **21**, 113–286 (1966).

Diversity of haematopoietic stem cell growth from a uniform population of cells

THE stochastic model for haematopoietic stem cell differentiation¹ with the addition of a maturation pathway and a 'resting' G_0 phase provides unifying explanations for three characteristics of the growth of colonies developing in the spleens of irradiated mice that have been injected with haematopoietic cells (Fig. 1).

First, it predicts that a homogeneous cell population can produce a high proportion of visible colonies that disappear between the 7th and 12th day of colony growth, with a nearly equal number of cells appearing during this interval; however, this requires the proviso that the self-renewal probability of colony-forming cells does not increase from the normal steady-state value until a few days after the irradiation and injection of haematopoietic cells. Second, the tiny transient erythroid colonies (devoid of colony-forming cells) observed in erythropoietically stimulated irradiated mice² can be predicted if there is a reduction in the threshold size for visibility, due to the increased contrast of more mature erythroid cells. Thus, these colonies may be produced by the same homogeneous population of multipotential cells producing the late-disappearing and late-

appearing colonies, but such colonies are those which run out of stem cells early, 60% predicted theoretically³, and so are not usually seen. Third, the late appearance of colonies (with higher colony-forming content) observed after injection of bone marrow cells from donor mice that have been treated with the anticancer agent 5-fluorouracil⁴ can be predicted by a 15% reduction in the number of maturing cell generations.

Physical cell separation shows unambiguously that spleen colony-forming cells are heterogeneous, but does not argue forcibly for the existence of a separate subpopulation of committed spleen colony-forming cells comparable in number to the multipotential population. We therefore suggest that the experimental results of Magli *et al.*⁵ fail to provide satisfactory evidence for the widely accepted claim "that the spleen colony method measures pluripotent stem cells only where macroscopic colonies are scored 11 days or later". Moreover, up to 80% of 10-day colonies contain colony-forming cells⁶. Thus, there is no compelling reason for scoring colonies later than 10 days, with the attendant increase in such technical problems as fewer countable colonies, linearity of colony counts with number of cells injected due to their larger size, possible migration from other sites, appearance of endogenous colonies and

more frequent difficulty with animal survival. Moreover, these data do not necessarily require a reassessment of work carried out over the past 20 years using this important assay for haematopoietic stem cells.

This highly oversimplistic model, with quite uncontroversial assumptions, demonstrates that results previously explained by postulating separate stem cell populations or selective drug sensitivity can be explained elegantly by a homogeneous population responding in different ways. Unless the minimum number of assumptions required to explain particular experimental data is found, it is unlikely that we will acquire a better understanding of the nature and regulation of haematopoietic stem cells. Our unifying model reverses the tendency to postulate a new cell population each time a new type of experimental result is observed. Elsewhere it is shown that these conclusions support the specific mechanism for the regulation of the number of haematopoietic stem cells proposed previously⁷, that they provide a straightforward explanation for the 'puzzle' concerning stem cell regeneration from cells with impaired self-renewal capacity⁸ and that the equation³ for calculating self-renewal has been misapplied.

N.M.B. is temporarily employed by the 'Association pour la Recherche sur le Cancer' and E.N. is holder of an INSERM Fellowship.

N. M. BLACKETT
E. NECAS
E. FRINDEL

Laboratoire de Cinétique
Cellulaire (U250),
Institut Gustave Roussy,
Villejuif, France

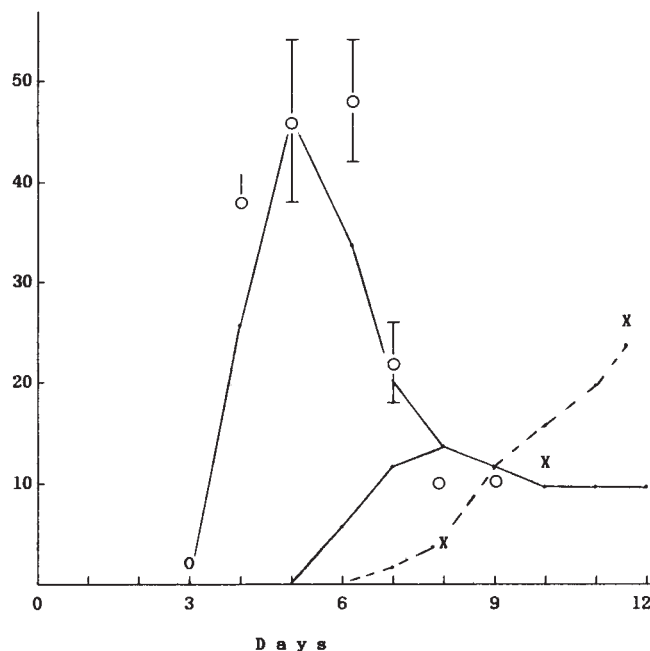
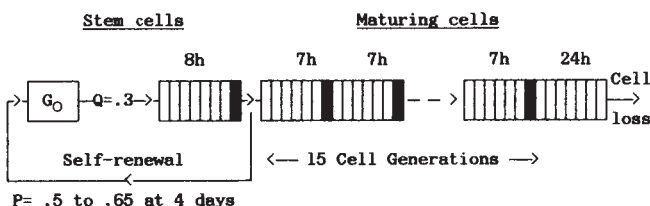


Fig. 1 Model with parameters predicting 40% loss of 7-day colonies and the number of colonies visible during growth. *a*, normal bone marrow (●); *b*, a reduction in the number of cell generations from 15 to 13 compared with experimental results, × ref. 4; *c*, a threefold reduction in the diameter threshold for colony visibility compared with experimental results, ○ ref. 2.

1. Till, J. E., McCulloch, E. A. & Siminovich, L. *Proc. natn. Acad. Sci. U.S.A.* **51**, 29-36 (1964).
2. Gregory, C. J., McCulloch, E. A. & Till, J. E. *J. cell. Physiol.* **86**, 1-8 (1975).
3. Vogel, H., Niewisch, H. & Maitoli, G. *J. cell. Physiol.* **72**, 221-228 (1968).
4. Hodgson, G. S. & Bradley, T. R. *Nature* **281**, 381-382 (1979).
5. Magli, M. C., Iscove, N. N. & Odartchenko, N. *Nature* **295**, 527-529 (1982).
6. Dumenil, D., Sainteny, F. & Frindel, E. *Leukemia Res.* **6**, 753-760 (1982).
7. Blackett, N. M. & Botnick, L. E. *Blood Cells* **7**, 417-423 (1981).
8. Schofield, R. & Lajtha, L. G. *Blood Cells* **9**, 467-473 (1983).

ISCOVE ET AL. REPLY—Blackett, Necas and Frindel show that our observations¹ of transient and late-appearing spleen colonies can be simulated by assuming an initially homogeneous colony-forming cell population and making an *ad hoc* adjustment of self-renewal probability. They conclude that our data do not prove that only the later colonies arise from pluripotent cells capable of extensive proliferation, and therefore claim that "there is no compelling reason for scoring colonies later than 10 days ...".

The essential point of our report was that early colonies are transient, and that the later-appearing multilineage colonies which contain primitive precursors are not derived from the early colonies. Therefore, the character of later colonies could no longer be invoked as proof of the primitiveness of the cells originating the early colonies. Faced with transient early colonies that had no primitive cells and contained practically only terminally differentiating erythroid cells, and now lacking any evidence for derivation from pluripotent precursors, we proposed the straightforward interpretation: that early and late colonies arise from different kinds of cell, and that the early transient colonies might in fact arise from committed precursors possessing limited proliferative potential.

There is an important additional consequence of the unrelatedness of early and late colonies which concerns the question of probability of 'self renewal'. The concept that this probability could increase during the development of spleen colonies was originally invoked to explain their increasing relative content of primitive precursors with time. An increase in self renewal probability during the course of spleen colony growth is the central assumption in the simulation performed by Blackett *et al.* However, our result removed both the need for this postulate and the evidence supporting it.

The important issue now is whether both early and late spleen colonies arise from an initially homogeneous precursor cell population. There is now abundant evidence that they do not: (1) Eight-day spleen colony-forming units (CFU-S[8]) are killed by 5-fluorouracil, while CFU-S[12] are relatively spared². (2) In sublethally irradiated mice, CFU-S[7] initiate DNA synthesis whereas CFU-S[11] do not³. (3) CFU-S[7] are twice as sensitive as CFU-S[11] to photoinactivation by the membrane-binding agent merocyanine 540 (ref. 4). (4) CFU-S[7] are resistant to inactivation by anti-Qa-m2 antibody plus complement, while most CFU-S[12] are inactivated⁵. (5) CFU-S sorted on the basis of binding of pokeweed mitogen segregate into a population high in proliferative capacity and ability to generate primitive precursors, and a second population having lower proliferative capacity and relatively enriched in CFU-S[7] (ref. 6). Sorting on the basis of H33342 staining⁷ and anti-H-2K binding similarly segregates cells forming early and late spleen colonies.

These experiments show conclusively that early and late spleen colonies are derived from different and separable types of precursor cells, and directly refute the model proposed by Blackett *et al.* Until contrary evidence appears, the available negative evidence provides more than adequate reason to question the notion

that spleen colonies scored at early times arise from pluripotent precursors with extensive proliferative potential.

N. N. ISCOVE

European Molecular Biology Laboratory,
Postbox 10 22 09,
D-6900 Heidelberg 1, FRG

M. C. MAGLI

Ontario Cancer Institute,
500 Sherbourne Street,
Toronto, Canada M4X 1K9

N. ODARTCHENKO

Swiss Cancer Institute,
CH-1066 Epalinges sur Lausanne,
Switzerland

1. Magli, M. C., Iscove, N. N. & Odartchenko, N. *Nature* **295**, 527-529 (1982).
2. Hodgson, G. S. & Bradley, T. R. *Nature* **281**, 381-382 (1979).
3. Chertkov, J. L. & Drize, N. J. *Cell Tissue Kinet.* **17**, 247-252 (1984).
4. Meagher, R. C. *et al.* *J. cell. Physiol.* **116**, 118-124 (1983).
5. Harris, R. A. *et al.* *Nature* **307**, 638-641 (1984).
6. Johnson, G. R. & Nicola, N. A. *J. cell. Physiol.* **118**, 45-52 (1984).
7. Baines, P. & Visser, J. W. M. *Expl. Hemat.* **11**, 701-708 (1983).
8. Visser, J. W. M. *et al.* *J. exp. Med.* **59**, 1576-1590 (1984).

DNA fingerprint analysis in immigration test-cases

RECENTLY, Jeffreys *et al.*¹ used 'fingerprints' of mini-satellites to check maternity in an immigration case. I do not dispute their conclusion that the boy (X) seeking immigration was the son of the putative mother (M), and not the son of an unrelated woman or of a sister of M, but their statistical analysis includes some deficiencies.

A complete analysis of the pedigree, which involved X, M and three undisputed offspring (B, S1 and S2) of M, requires construction of a series of likelihoods taking account of the distribution of DNA bands in each individual, for example information about homozygosity of particular mini-satellites is provided by whether or not all the sibs shared any particular band. It is useful first, however, to reanalyse the summary data given by Jeffreys *et al.*¹ in which, having assumed that X had the same father as B, S1 and S2, they deduced that 25 fragments, all of which were present in M, must have been inherited from X's mother. Following Jeffreys *et al.*, I assume that shared bands represent identical alleles at a hypervariable locus, and that the frequency (x) is the same for all bands. For the model where M is the aunt of X, I have computed the probability that M has the band and X has inherited his from a sister of M (Table 1). The relevant likelihood ratio is $L_a/L_m = 0.63^{25} \approx 10^{-5}$. Jeffreys *et al.*¹ computed, instead, the probability that M and a sister (assumed, but not stated, not to be an identical twin) share a band, which is not relevant because the band has to be

transmitted to X. Their value of 0.62 is so close to 0.63, however, that overall probabilities are little different.

The basic data on the 95 bands observed and the computation of likelihood in terms of the p_{ijk} , the prior probabilities of occurrence of the bands in different family members, are given in Table 2. Computation of p_{ijk} for each model is straightforward, but tedious.

For models unrelated (u) and mother (m) all possible genotypes of mother and father have to be considered. For example, if both are heterozygotes for a band, the prior frequency of the mating is $4q^2(1-q)^2$, $i=1$ since M has the band, and the conditional probabilities that $k=0, 1, 2$ and 3 are $1/64, 9/64, 27/64$ and $27/64$, respectively; the conditional probability that $j=0$ is $1/2 - q/2$ under model u and $1/4$ under model m. Summation over combinations of genotypes gives the p_{ijk} .

For model a, where M is the aunt, a two-stage process can be used. First, all possible matings among the parents of M and her sister and for these the conditional probabilities of each genotype of M and of the band inherited by X from his mother (that is, M's sister) have to be enumerated. For example, if both the grandparents are heterozygotes, the joint probability that M is a heterozygote and X does not inherit the band from his mother is $1/4$. From these data, prior probabilities of each of the possible inheritance patterns can be computed as a function of q . Second, these alternatives have to be taken with each possible genotype of the father, and the conditional probabilities that X has the band and the number of sibs with the band computed. For example, if r_3 denotes the prior probability that M is a heterozygote and X does not inherit the band from his mother, the prior probability of the corresponding mating of M to a heterozygote is $2q(1-q)r_3$, $i=1$ (M is a heterozygote), $j=0$ with probability $1/2$, and $k=0$ with probability $1/64$. Subsequent summation yields the p_{ijk} .

The likelihoods can be obtained for any value of q , but as there is no prior knowledge of band frequency in the Ghanaian population, q has to be estimated for each model by maximizing the likelihood. The important likelihood ratio is $L_a/L_m = 0.000035$ (Table 3) which, though six times higher than the probability ratio calculated by Jeffreys *et al.*, is still very small. If the likelihood computations are repeated with $q=0.14$, the value previously used¹, the corresponding ratio is much reduced (1.3×10^{-7}); and 0.14 was at the upper end of frequencies observed in British caucasians². Also, because frequencies of bands differ², these likelihood calculations overestimate L_a/L_m if M is indeed the mother.

There are many assumptions in these analyses, but there seems no justification in avoiding a formal method whatever the

Table 1 Probability that M carries the band and X has a maternally derived band for each model in the reduced analysis

Model	Probability	Likelihood ratio
M unrelated to X (u)	qx	$L_u/L_m = 0.26^{25} = 2.4 \times 10^{-15}$
M aunt of X (a)	$q(1+x)/2$	$L_u/L_a = 0.413^{25} = 2.5 \times 10^{-10}$
M mother of X (m)	q	$L_a/L_m = 0.63^{25} = 9.6 \times 10^{-6}$

x , Band frequency, assumed to be the same for each band ($x=0.26$); q , allele frequency ($x=2q-q^2$).

Table 2 The number of bands (n_{ijk}) present in the sample with $i=(0)1$ if the band is (absent) present in M, $j=(0)1$ if the band is (absent) present in X, $k=0, 1, 2, 3$ is the number of sibs with the band, and computation of likelihood

No. of sibs	Band absent in X ($j=0$)				Band present in X ($j=1$)			
	0	1	2	3	0	1	2	3
Band absent in M ($i=0$)	—	10	4	0	0	7	9	2
Band present in M ($i=1$)	2	5	8	4	3	13	13	15

p_{ijk} , Probability corresponding to n_{ijk} . The log likelihood is computed conditional on the occurrence of the band in at least one individual, with summation excluding the 000 class: $\ln L = \sum_i \sum_j \sum_k n_{ijk} [\ln p_{ijk} - \ln p_{000}]$.

Table 3 Maximum likelihood estimates of allele frequency (q_{max}) for each model, together with corresponding log likelihoods ($\ln L$) and likelihood ratios in the full analysis, assuming B, S1, S2 and X all have the same father

Model	q_{max}	$\ln L$	Likelihood ratios
u	0.32	-256.78	$L_u/L_m = 2.1 \times 10^{-9}$
a	0.26	-247.05	$L_u/L_a = 5.9 \times 10^{-5}$
m	0.18	-236.78	$L_a/L_m = 3.5 \times 10^{-5}$

model. As more information about the frequencies and inheritance of the mini-satellites becomes available, the models can be refined and the obvious power of the DNA-fingerprint technique utilized. Nevertheless, I do not believe that comparisons such as these give positive identification as suggested¹, only evidence that alternatives are highly unlikely.

I thank John Brookfield for providing the basic data, Susan Brotherstone for computational assistance, and members of Leicester University's Genetics Department for financial assistance.

WILLIAM G. HILL
Institute of Animal Genetics
University of Edinburgh,
West Mains Road
Edinburgh EH9 3JN, UK

1. Jeffreys, A. J., Brookfield, J. F. Y. & Semeonoff, R. *Nature* 317, 818-819 (1985).

2. Jeffreys, A. J., Wilson, V. & Thein, S. L. *Nature* 316, 76-79 (1985).

JEFFREYS ET AL. REPLY—We thank Professor Hill for his rigorous statistical analysis of the DNA fingerprint data of this immigration test-case. We are pleased that the new analysis gives ranges of values

for the likelihood ratios which are broadly similar to ours. We reiterate that even the new analysis includes a number of assumptions. The maximum likelihood values of q are based on only one family, and will be inaccurate. It is assumed that there is no variation in q between bands, and it is assumed that bands with the same mobility are always allelic and that bands with different mobilities never are.

We admit that we have not provided positive identification of the mother of X in a mathematical sense, since no probability calculation could do this. However, we believe that the probability of error in this case is so low that it would be legally appropriate to treat the test as providing positive identification.

A. J. JEFFREYS
J. F. Y. BROOKFIELD
R. SEMEONOFF

Department of Genetics,
University of Leicester,
Leicester LE1 7RH, UK

X;autosome translocations in females with Duchenne or Becker muscular dystrophy

RAY *et al.*¹ recently documented their efforts at isolating the gene for Duchenne muscular dystrophy by studying DNA cloned from the breakpoint of an X;21 translocation associated with Duchenne muscular dystrophy. However, the female patient² from whom this X-chromosomal material was derived³, clearly had Becker muscular dystrophy and not Duchenne dystrophy, as she was 20 years old and still ambulant. It thus seemed timely to critically review all the reported cases of

muscular dystrophy in females associated with an X;autosome translocation (Table 1) in order to try and designate them as Duchenne or Becker type.

Duchenne muscular dystrophy is an X-linked condition with onset in early childhood and a rapidly progressive course, so that the vast majority of affected boys lose the ability to walk by the age of 12 years⁴; Becker muscular dystrophy has an identical pattern of muscle involvement but follows a more benign course, and affected males will remain ambulant beyond 16 years of age and usually into adult life⁵. Inevitably there will also be occasional cases who may fall in the grey zone between the two and be a little milder than the usual Duchenne but not quite as mild as the usual Becker type and lose ability to walk between 13 and 16 years.

Careful attention has been paid to the diagnosis of male cases of either Duchenne or Becker dystrophy in the application of restriction fragment length polymorphism (RFLP) studies and detection of deletions. It would seem logical to pay the same attention to the initial diagnosis in the X;autosome translocation females from which the DNA sequences are cloned, rather than loosely referring to them all as Duchenne muscular dystrophy, as seems to have been the current practice.

Of the twelve cases documented to date, four⁶⁻⁹ conform closely to Duchenne dystrophy and a further four¹⁰⁻¹³ also seem to have a Duchenne severity but are still too young to be certain. One additional case¹⁴ is probably a Duchenne but was only published in abstract form with no mention of the age, which had to be surmised from the date of birth and timing of the abstract (approximately 14 years); she was said to be "unable to walk without assistance", which is also difficult to interpret. One case² conforms to Becker type and another¹⁵ also seems likely to be, but is too young (13 years) to be certain about ambulation beyond 16 years. The remaining case¹⁶ (published only in abstract form) has insufficient clinical data to draw any conclusions. Two of these cases are somewhat atypical, one¹⁰ having an associated dysmorphic syndrome and the other⁹ an associated Turner's syndrome.

Recent studies¹⁷⁻²⁰ have shown that the gene locus for Becker muscular dystrophy is very close to that for Duchenne and it seems likely that the two conditions are allelic. Studies with some of the probes (for example pERT 87.8)²¹ have shown deletions in approximately 9% (5/57) of cases of Duchenne muscular dystrophy but none with cases of Becker muscular dystrophy.

In the report of Ray *et al.*¹, only one Duchenne dystrophy patient out of 50 studied showed a deletion for their XJ1 clone, and this was also the only case which showed a deletion for pERT 87. In

Table 1 X;autosome translocations in girls with muscular dystrophy

Reference	Autosomal breakpoint	Age at reporting	Age of onset	Loss of ambulation	Still walking	IQ	Conclusions
Canki <i>et al.</i> (1979) ¹⁰	3q13	4 yr	10 mth	—	+	R	? DMD*
Lindenbaum <i>et al.</i> (1979) ⁶	1p34	8 yr	<2 yr	+(8 yr)	—	N	DMD†
Greenstein <i>et al.</i> (1980) ¹⁶	11q13	16 yr	?	?	?	?	? DMD or BMD
Jacobs <i>et al.</i> (1981) ⁷	5q35	9 yr	4 yr	—	+	N	DMD
Zatz <i>et al.</i> (1981) ⁸	6q21	11 yr	5 yr	+(10 yr)	—	N	DMD
Emanuel <i>et al.</i> (1983) ¹¹	9p22	9 yr	<2 yr	—	+	R	? DMD
Nielsen <i>et al.</i> (1983) ¹⁵	11q23	13 yr	<2 yr	—	+	N	? BMD
Perez Vidal <i>et al.</i> (1983) ¹²	6q16	4½ yr	<3 yr	—	+	?	? DMD
MacLeod <i>et al.</i> (1983) ¹⁴	2q36	?14 yr	<2 yr	+	—	R	? DMD
Verellen-Dumoulin <i>et al.</i> (1984) ²	21p12	20 yr	2 yr	—	+	N	BMD
Bjerglund-Nielsen (1984) ⁹	9p21	23 yr‡	<2 yr	+(12 yr)	—	R	DMD§
Saito <i>et al.</i> (1985) ¹³	4q26	3 yr	2 yr	—	+	?	? DMD

N, normal; R, retarded; DMD, Duchenne muscular dystrophy; BMD, Becker muscular dystrophy.

* Associated dysmorphic syndrome.

† Personally examined.

‡ Died.

§ Associated Turner's syndrome.

addition, all five patients of Monaco *et al.*²¹ with deletions for pERT87 were found to have deletions for XJ-1.

Although all the reported cases of X;autosome translocation with muscular dystrophy have apparently had the breakpoint in the same region of the short arm of the X chromosome (Xp21), there may be variations from case to case in the exact location of the breakpoint²². It may thus be of interest and importance when cloning sequences from these X-chromosome breakpoint regions, and applying them in the detection of RFLPs close to the Duchenne/Becker dystrophy genes, and looking for deletions in individual male cases, to know whether the original source came from a Duchenne or a Becker type case, and in addition to have adequate documentation of the dystrophic nature of the underlying muscle disorder.

Note added in proof: A further case, conforming to Duchenne severity, with an X;5 translocation, has recently been documented by Nevin *et al.*²³.

VICTOR DUBOWITZ

Department of Paediatrics &
Neonatal Medicine and
Jerry Lewis Muscle Research Centre,
Royal Postgraduate Medical School,
Hammersmith Hospital,
Du Cane Road,
London W12 0HS, UK

12. Perez Vidal, M. T. *et al.* *Rev. Neurol. (Barcelona)* **XL**, 51, 155–158 (1983).
13. Saito, F. *et al.* *Hum. Genet.* **71**, 370–371 (1985).
14. MacLeod, P. M., Holden, J. & Masotti, R. *Am. J. hum. Genet.* **35**, 104A (1983).
15. Nielsen, L. B. *et al.* *Clin. Genet.* **23**, 242 (1983).
16. Greenstein, R. M. *et al.* *Cytogenet. Cell Genet.* **27**, 268 (1980).
17. Kingston, H. M. *et al.* *J. med. Genet.* **20**, 255–258 (1983).
18. Kingston, H. M. *et al.* *Hum. Genet.* **67**, 6–17 (1984).
19. Brown, C. S. *et al.* *Hum. Genet.* **71**, 62–74 (1985).
20. Dorkins, H. *et al.* *Hum. Genet.* **71**, 103–107 (1985).
21. Monaco, A. P. *et al.* *Nature* **316**, 842–845 (1985).
22. Boyd, Y. & Buckle, V. J. *Clin. Genet.* **29**, 108–115 (1986).
23. Nevin, N. G. *et al.* *J. med. Genet.* **23**, 171–173 (1986).

WORTON REPLIES—Dubowitz raises a very valid concern about the phenotypes of girls with myopathic disease secondary to an X;autosome translocation. Certainly the X;21 translocation patient whose junction has been cloned has a mild form of muscular dystrophy more like the Becker form of the disease and should perhaps be reclassified as such.

However, there are two very important points to be made about the translocation patients in general. First, all have translocation exchange points in band Xp21 despite the fact that some have a severe (Duchenne) phenotype and others appear to have the milder (Becker) phenotype. This fits with the linkage data that map the gene for both Becker and Duchenne muscular dystrophy to this region of the X chromosome. As Dubowitz points out the two diseases may in fact be due to allelic mutations at the same locus. Thus, although it is important to have the phenotype of patients accurately described, the Duchenne and Becker variants of muscular dystrophy may simply represent different manifestations of the same disease.

The second important point, and one not discussed by Dubowitz, is that the expression of muscle disease in these girls who are heterozygous for the mutation (translocation) is a function of the non-random inactivation of the normal X chromosome carrying the wild-type allele. In many of the girls examined, the prefer-

ence for normal X inactivation is not complete; up to 10% of the lymphocytes examined displayed a late replicating (inactive) translocation X and an early replicating (active) normal X. If this pattern holds in muscle, a proportion of the nuclei in a multi-nucleated myotube may be capable of producing the normal gene product. Since the 50% level of gene product expected of a non-translocation carrier female is sufficient to prevent the disease, a level of 5% or 10% may well be sufficient to modify the severity of the disease. In translocation patients a mutation at the 'Duchenne gene', therefore, may result in a Becker phenotype. This complicates the picture and suggests that neither term is really appropriate to the translocation females. Only through detailed genetic studies of the Becker/Duchenne muscular dystrophy locus will the true relationship between the two diseases be understood.

R. G. WORTON

The Hospital for Sick Children,
555 University Avenue,
Toronto, Ontario,
Canada M5G 1X8

Matters Arising

Matters Arising is meant as a vehicle for comment and discussion about papers that appear in *Nature*. The originator of a Matters Arising contribution should initially send his manuscript to the author of the original paper and both parties should, wherever possible, agree on what is to be submitted. Neither contribution nor reply (if one is necessary) should be longer than 500 words and the briefest of replies, to the effect that a point is taken, should be considered.

1. Ray, P. N. *et al.* *Nature* **318**, 672–675 (1985).
2. Verellen-Dumoulin, C. *et al.* *Hum. Genet.* **67**, 115–119 (1984).
3. Worton, R. G. *et al.* *Science* **224**, 1447–1449 (1984).
4. Dubowitz, V. *Muscle Disorders in Childhood* (Saunders, London, 1978).
5. Emery, A. E. H. & Skinner, R. *Clin. Genet.* **10**, 189–201 (1976).
6. Lindenbaum, R. H. *et al.* *J. med. Genet.* **16**, 389–392 (1979).
7. Jacobs, P. A. *et al.* *Am. J. hum. Genet.* **33**, 513–518 (1981).
8. Zatz, M. *et al.* *J. med. Genet.* **18**, 442–447 (1981).
9. Bjerglund-Nielsen, L. & Nielsen, I. M. *Ann. Génét.* **27**, 173–177 (1984).
10. Canki, N., Dutrillaux, B. & Tivadar, I. *Ann. Génét.* **22**, 35–39 (1979).
11. Emanuel, B. S., Zackai, E. H. & Tucker, S. H. *J. med. Genet.* **20**, 461–463 (1983).